



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 174319

TO: Patricia Duffy
Location: REM-3B05/3C18
Art Unit: 1645
Wednesday, December 28, 2005

Case Serial Number: 09/767041

From: Paul Schulwitz
Location: Biotech-Chem Library
REM-1A65
Phone: 571-272-2527

Paul.schulwitz@uspto.gov

Search Notes

Examiner Duffy,

Please review the attached search results.

If you have any questions or if you would like to refine the search query, please feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz
Technical Information Specialist
REM-1A65
571-272-2527

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

From: Duffy, Patricia
Sent: Thursday, December 15, 2005 8:56 AM
To: STIC-Biotech/ChemLib
Subject: SEQUENCE SEARCH 09/767,041

IN RE:09/767,041

PLEASE SEARCH SEQ ID NO:9 AND OLIGOMERS THEREOF.

Patricia A. Duffy, Ph.D.
Art Unit 1645
Remsen 3B05; Mailbox 3C18
571-272-0855

RECEIVED
DEC 15 2005
STIC/BIOTECH DIVISION
(STIC)

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____



THIS PAGE BLANK (USPTO)

Schulwitz, Paul

From: Pak, Michael
Sent: Wednesday, December 21, 2005 11:38 AM
To: Schulwitz, Paul
Cc: Duffy, Patricia
Subject: RE:

Dear Paul or STIC,

Please search the extra long sequence. The search is necessary for the examination of the application.

Thanks,

Mike Pak

-----Original Message-----

From: Duffy, Patricia
Sent: Wednesday, December 21, 2005 11:34 AM
To: Pak, Michael
Cc: Schulwitz, Paul
Subject:
Importance: High

Dear Mike,

Please approve the search for US 09/767,041. The nucleic acid is 6000 bp long and the claim requires that the bacterium have the entire sequence or specific oligomeric fragments thereof. The search is going to take over 20 hrs and Paul Schulwitz needs your approval to run this search. Please send approval to Paul to expedite.

Thanks. Mucho have a great holiday.

Pat Duffy

Dear Stic:
In re: 09/767,041

Please search SEQ ID NO:9 and oligomers thereof.

Patricia A. Duffy, Ph.D.
Art Unit 1645
Remsen 3B05; Mailbox 3C18
571-272-0855

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 03:12:19 ; Search time 22525 Seconds

(without alignments)
17644.813 Million cell updates/sec

Title: US-09-767-041-9

Perfect score: 6992

Sequence: 1 atcgccaacgaagattgca.....gcacaacgcaaaagaagctt 6992

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_to:*
10: gb_tes:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6992	100.0	6992	1	AF155804 Streptococcus
2	1195.8	17.1	13401	1	AF118389 Streptococcus
3	955.6	13.7	17468	1	AY057915 Streptococcus
4	955.6	13.7	17468	6	AX283715 Streptococcus
5	892.6	12.8	21562	1	STH289861 Streptococcus
6	892.6	12.8	110000	1	CP000023_09 Streptococcus
7	889	12.7	16032	6	CQ963297 Streptococcus
8	888.8	12.7	19918	1	CR931662 Streptococcus
9	885.6	12.7	14943	1	SPC93162 Streptococcus
10	869.4	12.4	22405	1	CR931663 Streptococcus
11	866.2	12.4	18517	1	CR931663 Streptococcus
12	865.2	12.4	17717	1	CR931661 Streptococcus
13	851.8	12.2	18634	1	CR931664 Streptococcus
14	851.8	12.2	18626	1	CR931665 Streptococcus
15	851.6	12.2	16926	1	CR931657 Streptococcus
16	848.6	12.1	17082	1	CR931654 Streptococcus
17	848.6	12.1	16532	1	CR931655 Streptococcus
18	841.8	12.0	17948	1	CR931653 Streptococcus

19	840.2	12.0	17213	1	CR931656 Streptococcus
20	839.6	12.0	14202	1	AY376403 Streptococcus
21	798.8	11.4	18239	1	AF349539 Streptococcus
22	798.8	11.4	21365	1	AE014245 Streptococcus
23	798.8	11.4	34980	6	CQ655070 Streptococcus
24	798.8	11.4	34980	6	AX954530 Streptococcus
25	743.4	10.6	17596	1	AF355776 Streptococcus
26	729.4	10.4	4513	1	AF332894 Streptococcus
27	729.4	10.4	4513	1	AF363060 Streptococcus
28	729.4	10.4	4522	1	AF332914 Streptococcus
29	729.4	10.4	4888	1	AF332893 Streptococcus
30	729.4	10.4	4888	1	AF363058 Streptococcus
31	727.8	10.4	4888	1	AF363057 Streptococcus
32	727.8	10.4	4888	1	AF363059 Streptococcus
33	727.8	10.4	6865	6	BD166249 Streptococcus
34	727.8	10.4	6865	6	AR613460 Streptococcus
35	727.8	10.4	9987	1	AB050723 Streptococcus
36	726.8	10.4	4768	1	AF332896 Streptococcus
37	726.8	10.4	4768	1	AF363055 Streptococcus
38	726.8	10.4	4768	1	AF363056 Streptococcus
39	726.8	10.4	17276	1	AF163833 Streptococcus
40	725.2	10.4	4411	1	AF332897 Streptococcus
41	725.2	10.4	4411	1	AF332898 Streptococcus
42	725.2	10.4	4411	1	AF332899 Streptococcus
43	725.2	10.4	4411	1	AF332900 Streptococcus
44	725.2	10.4	95596	6	AX602206 Streptococcus
45	725.2	10.4	128050	1	SAG766849 Streptococcus

ALIGNMENTS

RESULT 1
LOCUS AF155804
DEFINITION Streptococcus suis strain 655 CpsII (cpsII) gene, partial cds;
CpsII (cpsII) genes, complete cds; and CpsII (cpsII) gene, partial
CDS.
ACCESSION AF155804
VERSION AF155804.1 GI:6601338
KEYWORDS Streptococcus suis
SOURCE Streptococcus suis
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 6992)
AUTHORS Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F.,
Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
TITLE Identification and characterization of the cps locus of
Streptococcus suis serotype 2: the capsule protects against
phagocytosis and is an important virulence factor
JOURNAL Infect. Immun. 67 (4), 1750-1756 (1999)
PUBMED 10085014
JOURNAL PUMED 10085014
AUTHORS 2 (bases 1 to 6992)
TITLE Smith,H.E., Veenbergen,V., van der Velde,J., Damman,M.,
Wisselink,H.J. and Smits,M.A.
JOURNAL The cps genes of Streptococcus suis serotypes 1, 2, and 9:
development of rapid serotype-specific PCR assays
JOURNAL J. Clin. Microbiol. 37 (10), 3146-3152 (1999)
PUBMED 10488168
REFERENCE 3 (bases 1 to 6992)
AUTHORS Smith,H.E., Veenbergen,V., Van der Velde,J., Damman,M.,
Wisselink,H.J. and Smits,M.A.
TITLE Direct Subtyping
JOURNAL Submitted (02-JUN-1999) Bacteriology, Animal Science and Health,
P.O. Box 65, Lelystad 8200 AB, The Netherlands
FEATURES
source
1..6992
/organism="Streptococcus suis"
/mol_type="genomic DNA"
/strain="6555"
/serotype="1"

Db	421	ATATGCAAGTTTATTTGATCACAATAACAAATTCGTAATCTTGTGATGGTAG	480
Oy	481	TTTTAGGTACGAAATAGATAAAATTAATTAATCATTAACCGTCAATTAATCTGTGAAG	540
Db	481	TTTTTAGTACGAATAGATAAAATTAATTAATCATTAACCGTCAATTAATCTGTGAAG	540
Oy	541	AACTATATAGAGTTTCAACAAAGGGAAGTGCTGACCAAGTCTTTATTAATCTCAAGTGTG	600
Db	541	AAAGCTATAGAGTTTCAACAAAGGGAAGTGCTGACCAAGTCTTTATTAATCTCAAGTGTG	600
Oy	601	AGTTTTTGAAGTAAGCAATTGCTTGACAGTTTGAGTTGATGTAATGATATGAATGAACG	660
Db	601	AGTTTTTGAAGTAAGCAATTGCTTGACAGTTTGAGTTGATGTAATGATATGAATGAACG	660
Oy	661	TTGATATTAATTAATTCATTCGGTTTCTGCGTTGAAAAAATAAATTCACACTGTAAGTGTG	720
Db	661	TTGATATTAATTCATTCGGTTTCTGCGTTGAAAAAATAAATTCACACTGTAAGTGTG	720
Oy	721	ACCATAGCATTGTAACTTTTTCACAAATTTTATTAAGCTTAATCATATCATATGAAC	780
Db	721	ACCATAGCATTGTAACTTTTTCACAAATTTTATTAAGCTTAATCATATCATATGAAC	780
Oy	781	GACTTTTGATATTAATCTCGAGCGGTAGTCCGGTTAATTAATTTGCGTATAGTTCTAATTT	840
Db	781	GACTTTTGATATTAATCTCGAGCGGTAGTCCGGTTAATTAATTTGCGTATAGTTCTAATTT	840
Oy	841	TGTTAGTTTCCAAATTAATCTGTAGAGATGGTGACCGGCTAATTTTGTCTCAGAAACGAGTTG	900
Db	841	TGTTAGTTTCCAAATTAATCTGTAGAGATGGTGACCGGCTAATTTTGTCTCAGAAACGAGTTG	900
Oy	901	GACAGAAATGACGCAATATTAATTAATCAATTCGAAGTTTCGATCGATGTATGTATGCTGAAG	960
Db	901	GACAGAAATGACGCAATATTAATTAATCAATTCGAAGTTTCGATCGATGTATGTATGCTGAAG	960
Oy	961	AGCGCAAAAAAGACTTGTCTCAGCCCAAAACCAATGCAAGGTTGGGTATGTTTAAATAG	1020
Db	961	AGCGCAAAAAAGACTTGTCTCAGCCCAAAACCAATGCAAGGTTGGGTATGTTTAAATAG	1020
Oy	1021	GAAAAAGATCCTGAATTAATCTCCCAATTTGACATTTCTATGACAAAAACAAGTTTAAAGC	1080
Db	1021	GAAAAAGATCCTGAATTAATCTCCCAATTTGACATTTCTATGACAAAAACAAGTTTAAAGC	1080
Oy	1081	AGTTACCAACAGTTTATATGATGTTTAAATTTGGCGATATAGCTAGTTGGTACACGCTCAC	1140
Db	1081	AGTTACCAACAGTTTATATGATGTTTAAATTTGGCGATATAGCTAGTTGGTACACGCTCAC	1140
Oy	1141	CTACAGTTGATGAATTTGAAAAATATACTCTCGTGTCAAAAGACAGATTTGAGTTTAAAC	1200
Db	1141	CTACAGTTGATGAATTTGAAAAATATACTCTCGTGTCAAAAGACAGATTTGAGTTTAAAC	1200
Oy	1201	CAGGGAATTAACAGGCTCTCGGACAGTTAATGATCGTAATTAATCAACAATTCACAGACG	1260
Db	1201	CAGGGAATTAACAGGCTCTCGGACAGTTAATGATCGTAATTAATCAACAATTCACAGACG	1260
Oy	1261	TAGTTCCGTTGACCTTAGCATATCATTTGATTAATTTGGAATCTATCTGGTCAATATTAATAATTT	1320
Db	1261	TAGTTCCGTTGACCTTAGCATATCATTTGATTAATTTGGAATCTATCTGGTCAATATTAATAATTT	1320
Oy	1321	TATTAAGAACAAGTGAAGTTGTATTTGTGAAGAGGAAAGTAAAGTATATGAAG	1380
Db	1321	TATTAAGAACAAGTGAAGTTGTATTTGTGAAGAGGAAAGTAAAGTATATGAAG	1380
Oy	1381	TTTGTGTTGCTGCTTCTCAGGAGGACATTTGACATCTGTTATTTGTTAAACCGTTT	1440
Db	1381	TTTGTGTTGCTGCTTCTCAGGAGGACATTTGACATCTGTTATTTGTTAAACCGTTT	1440
Oy	1441	GGAAGGAAGAAGACGTTTGTGGGTAAACATTTGATTAAGAAGATGCAAGAAGCTCTTTGA	1500
Db	1441	GGAAGGAAGAAGACGTTTGTGGGTAAACATTTGATTAAGAAGATGCAAGAAGCTCTTTGA	1500
Oy	1501	AGATGAAAAAATGTATCATATTAATTTTCCAAACAATCGCAATCTCATTAATTTAGTGA	1560
Db	1501	AGATGAAAAAATGTATCATATTAATTTTCCAAACAATCGCAATCTCATTAATTTAGTGA	1560

Dp	1501	AGAAATGAAAAATGATACCATGTTACTTCCAAACAAATCCGAATCTCATTAATTAATGGA	1560
Qy	1561	AAAAATCTTTCTTAGCTTTCAAATTTTACGTATGGAACCAAGATTAATTTTCAT	1620
Dp	1561	AAAAATCTTTCTTAGCTTTCAAATTTTACGTATGGAACCAAGATTAATTTTCAT	1620
Qy	1621	CTGGTGGGCGCGTGTGCTGCCCTCTTTTACATCGGAAAACTATTTGGAGCAAGCGA	1680
Dp	1621	CTGGTGGGCGCGTGTGCTGCCCTCTTTTACATCGGAAAACTATTTGGAGCAAGCGA	1680
Qy	1681	TTTATATGAAGTATTTGATCGAGTATTAATCTATCATTTACTGAAAACTAGTTATTC	1740
Dp	1681	TTTATATGAAGTATTTGATCGAGTATTAATCTATCATTTACTGAAAACTAGTTATTC	1740
Qy	1741	CCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAATGAAGAGTATATCTTAATCTA	1800
Dp	1741	CCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAATGAAGAGTATATCTTAATCTA	1800
Qy	1801	TTAATCTGGGGAGTATTTTAAATGATTTTGTACAGTATAGAACTCATGAAACAAGTT	1860
Dp	1801	TTAATCTGGGGAGTATTTTAAATGATTTTGTACAGTATAGAACTCATGAAACAAGTT	1860
Qy	1861	TAATCGATGTGTAATAAGATGTGATTTATTTGAAAAAAAATGGAAGTATTAACCGACGAAT	1920
Dp	1861	TAATCGATGTGTAATAAGATGTGATTTATTTGAAAAAAAATGGAAGTATTAACCGACGAAT	1920
Qy	1921	ATTTATTTCAACAGAGATATTTCTGACTATATTTCCAAATATTTGCAAGTATAAAAATTTCT	1980
Dp	1921	ATTTATTTCAACAGAGATATTTCTGACTATATTTCCAAATATTTGCAAGTATAAAAATTTCT	1980
Qy	1981	CAGTTACAAAGAAATGGAACAATATTTAAACAAATCGAAGTATTTTCCACGAGAG	2040
Dp	1981	CAGTTACAAAGAAATGGAACAATATTTAAACAAATCGAAGTATTTTCCACGAGAG	2040
Qy	2041	CCCCGCTACTTTTATGAAATTCATATTCGAAAGAAAAAACAATATTTGTTTCTTAGACA	2100
Dp	2041	CCCCGCTACTTTTATGAAATTCATATTCGAAAGAAAAAACAATATTTGTTTCTTAGACA	2100
Qy	2101	AAAAAAGATGATGATGACATGTAATGATCATCAAGTATGTTGTGAAGAAATTTTACA	2160
Dp	2101	AAAAAAGATGATGATGACATGTAATGATCATCAAGTATGTTGTGAAGAAATTTTACA	2160
Qy	2161	AGATTAATATATTTTATTTATGAAAAATATGATGATTTGTTGAAAAAAATATTTGAAGT	2220
Dp	2161	AGATTAATATATTTTATTTATGAAAAATATGATGATTTGTTGAAAAAAATATTTGAAGT	2220
Qy	2221	TTCTTAACCAACTATCTTATACATCAATAATATTTTGTGGAAGATTTAAACAAT	2280
Dp	2221	TTCTTAACCAACTATCTTATACATCAATAATATTTTGTGGAAGATTTAAACAAT	2280
Qy	2281	AGTTGAAAAATTTATATGAGATCAAGAAATGAAATATTAATAAAAGATGATTTTGTA	2340
Dp	2281	AGTTGAAAAATTTATATGAGATCAAGAAATGAAATATTAATAAAAGATGATTTTGTA	2340
Qy	2341	TGGCTTATCATTAATTTTCTCAGATTTTACTGAGAGGATACAGATATTTATCATCTTCT	2400
Dp	2341	TGGCTTATCATTAATTTTCTCAGATTTTACTGAGAGGATACAGATATTTATCATCTTCT	2400
Qy	2401	CTCGAGAGATGCAACCATTAAGTTCTTCTCAGAAATACCTGTATATATTTTAAATATTC	2460
Dp	2401	CTCGAGAGAGATGCAACCATTAAGTTCTTCTCAGAAATACCTGTATATATTTTAAATATTC	2460
Qy	2461	TCAGAGATTTATATGTTGAATTTTACAAAAGATGAGCAAAATATTAAGAAAAATGAGATTA	2520
Dp	2461	TCAGAGATTTATATGTTGAATTTTACAAAAGATGAGCAAAATATTAAGAAAAATGAGATTA	2520
Qy	2521	TGAACGAGTTAAATGTTACAGATTAATTTCTTAATATATCGAAAAAAACCTATGATATATGT	2580
Dp	2521	TGAACGAGTTAAATGTTACAGATTAATTTCTTAATATATCGAAAAAAACCTATGATATATGT	2580
Qy	2581	ACGTGTTTGAATTTTATTAAGAAATGATCGAGCTTTGGAATACATATTTACCAAGATTTGT	2640
Dp	2581	ACGTGTTTGAATTTTATTAAGAAATGATCGAGCTTTGGAATACATATTTACCAAGATTTGT	2640

2641 GTTATTTGATAGATPAAAAAACATGGCTAAGATTAAGATTGGTCTAATTGGGTTTCG 2700
2641 GTTATTTGATAGATPAAAAAACATGGCTAAGATTAAGATTGGTCTAATTGGGTTTCG 2700
2701 CTTCACATGATTTTGTGGCAATCTTTTATCAATGAACGAAACAGCTTATTTATTT 2760
2701 CTTCACATGATTTTGTGGCAATCTTTTATCAATGAACGAAACAGCTTATTTATTT 2760
2701 CTTCACATGATTTTGTGGCAATCTTTTATCAATGAACGAAACAGCTTATTTATTT 2760
2761 AAGTATCTAAATGTCAGATGACCTTATTCAGACAAATTAAGAAAAATATGAATTT 2820
2761 AAGTATCTAAATGTCAGATGACCTTATTCAGACAAATTAAGAAAAATATGAATTT 2820
2761 AAGTATCTAAATGTCAGATGACCTTATTCAGACAAATTAAGAAAAATATGAATTT 2820
2821 TCAAAATAGATTAATCTAAATATGGAATTAAGATATTAAGGAAAAATCAACATCT 2880
2821 TCAAAATAGATTAATCTAAATATGGAATTAAGATATTAAGGAAAAATCAACATCT 2880
2881 TCTCCTATTTGCTTTACAGATGATCTATTTGATGAATTCCTAAATGCAAGAAATTTAGT 2940
2881 TCTCCTATTTGCTTTACAGATGATCTATTTGATGAATTCCTAAATGCAAGAAATTTAGT 2940
2941 TTTTATTTTGTGTAAGAAAGTTAAAAATGAAAAATTAATTTTAAAGAAATTTACT 3000
2941 TTTTATTTTGTGTAAGAAAGTTAAAAATGAAAAATTAATTTTAAAGAAATTTACT 3000
2941 TTTTATTTTGTGTAAGAAAGTTAAAAATGAAAAATTAATTTTAAAGAAATTTACT 3000
3001 AAAAAATTAATGTTGATTTTGTGAGAGATGATGATTAATTAATTTAATTTATGACC 3060
3001 AAAAAATTAATGTTGATTTTGTGAGAGATGATGATTAATTAATTTAATTTATGACC 3060
3061 CGGAATATTTTATTTTAAAGTACTCTGCTGATTAATTTTATTCAGACGAAAGATG 3120
3061 CGGAATATTTTATTTTAAAGTACTCTGCTGATTAATTTTATTCAGACGAAAGATG 3120
3061 CGGAATATTTTATTTTAAAGTACTCTGCTGATTAATTTTATTCAGACGAAAGATG 3120
3061 CGGAATATTTTATTTTAAAGTACTCTGCTGATTAATTTTATTCAGACGAAAGATG 3120
3121 TATTTTATTTAATTTTATGAAATTTAATTTTATTCATATAAATTTTGAAGAACTAACG 3180
3121 TATTTTATTTAATTTTATGAAATTTAATTTTATTCATATAAATTTTGAAGAACTAACG 3180
3121 TATTTTATTTAATTTTATGAAATTTAATTTTATTCATATAAATTTTGAAGAACTAACG 3180
3121 TATTTTATTTAATTTTATGAAATTTAATTTTATTCATATAAATTTTGAAGAACTAACG 3180
3181 TAATATTAATAAATGAAATTTTATGTTTATTAAGGCTAATATTAATTTTGTTCAG 3240
3181 TAATATTAATAAATGAAATTTTATGTTTATTAAGGCTAATATTAATTTTGTTCAG 3240
3241 TAGTCACAGATGATTTGTTGAAATTAATTTTGAAGATTAATTTGAGATTTTACTGCTC 3300
3241 TAGTCACAGATGATTTGTTGAAATTAATTTTGAAGATTAATTTGAGATTTTACTGCTC 3300
3301 CCATTAATTTGATTAATGCAATTAATTAATTAATTTGATTCATTAATTAATTAATGATT 3360
3301 CCATTAATTTGATTAATGCAATTAATTAATTAATTTGATTCATTAATTAATTAATGATT 3360
3361 AATAAATAATTAATAAATGATGATCTTTTATGTTTATTAAGTTTATGATTAATCTGAT 3420
3361 AATAAATAATTAATAAATGATGATCTTTTATGTTTATTAAGTTTATGATTAATCTGAT 3420
3361 AATAAATAATTAATAAATGATGATCTTTTATGTTTATTAAGTTTATGATTAATCTGAT 3420
3421 TGTATATTAATTAATAAATGGAAGATTAATTTTATGACAGACACCTTATAGGACTAG 3480
3421 TGTATATTAATTAATAAATGGAAGATTAATTTTATGACAGACACCTTATAGGACTAG 3480
3421 TGTATATTAATTAATAAATGGAAGATTAATTTTATGACAGACACCTTATAGGACTAG 3480
3481 ACTATCTTAATPAACGGCGTCAAAAAGGTTGGCTTTATGAACTATTCACGTTAA 3540
3481 ACTATCTTAATPAACGGCGTCAAAAAGGTTGGCTTTATGAACTATTCACGTTAA 3540
3541 ATACCACTACAAATTAATGTTCAATTCCTTATGACCTTATTAATAAATAAATAATG 3600
3541 ATACCACTACAAATTAATGTTCAATTCCTTATGACCTTATTAATAAATAAATAATG 3600
3601 AACAAATTTTCTTGTGCTGCTTATTAACGATCTAATTAAGTGAATGCAAGATG 3660
3601 AACAAATTTTCTTGTGCTGCTTATTAACGATCTAATTAAGTGAATGCAAGATG 3660
3601 AACAAATTTTCTTGTGCTGCTTATTAACGATCTAATTAAGTGAATGCAAGATG 3660
3661 GTAGTTTATGCTGACATATTAATTAATGCTTGTATGAGATTAATTAATGAGTGAAT 3720
3661 GTAGTTTATGCTGACATATTAATTAATGCTTGTATGAGATTAATTAATGAGTGAAT 3720

3721 TTGCTGATTAATAAAGCTAATAGTAAATTTGTAATCTACTATTAATTTAAATACG 3780
3721 TTGCTGATTAATAAAGCTAATAGTAAATTTGTAATCTACTATTAATTTAAATACG 3780
3781 AATGCTTATCAATGAATTTTGGCTTTTATTTATTTAGAGATCAAGTAAAGAGCTA 3840
3781 AATGCTTATCAATGAATTTTGGCTTTTATTTATTTAGAGATCAAGTAAAGAGCTA 3840
3781 AATGCTTATCAATGAATTTTGGCTTTTATTTATTTAGAGATCAAGTAAAGAGCTA 3840
3841 GATTTATTAATTTATCAAGAAATGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 3900
3841 GATTTATTAATTTATCAAGAAATGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 3900
3901 ATGGAATATCCGAATATTCAGTTAGGGAATCTGGCTGGAAGTATTCAGGCTATATAT 3960
3901 ATGGAATATCCGAATATTCAGTTAGGGAATCTGGCTGGAAGTATTCAGGCTATATAT 3960
3961 CATTTTATTAATTAATCAAGAAATGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 4020
3961 CATTTTATTAATTAATCAAGAAATGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 4020
3961 CATTTTATTAATTAATCAAGAAATGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 4020
4021 TTTAATAAATAAAGTTATGAGATTAATGGGAAACAGACTATTTTATTTACATCATG 4080
4021 TTTAATAAATAAAGTTATGAGATTAATGGGAAACAGACTATTTTATTTACATCATG 4080
4081 CCATATTTTCTATATGAAACATATGATCCGATTAATTAATTAATTAATTAATTAAT 4140
4081 CCATATTTTCTATATGAAACATATGATCCGATTAATTAATTAATTAATTAATTAAT 4140
4141 CTTCATATGATTTGGAATTAATTAATTAATTTTAAAGATTAAGAGCAAAATTAAT 4200
4141 CTTCATATGATTTGGAATTAATTAATTAATTTTAAAGATTAAGAGCAAAATTAAT 4200
4201 GATTTAATTTCAATTAATTTGATTAATTAATTTTAAATTTTAAATTTTAAATTTT 4260
4201 GATTTAATTTCAATTAATTTGATTAATTAATTTTAAATTTTAAATTTTAAATTTT 4260
4201 GATTTAATTTCAATTAATTTGATTAATTAATTTTAAATTTTAAATTTTAAATTTT 4260
4261 AACAGATTAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAAT 4320
4261 AACAGATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 4320
4321 ACTGATGATTTGGAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4380
4321 ACTGATGATTTGGAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4380
4381 TACAGAAATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4440
4381 TACAGAAATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4440
4441 GGTAAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTATTAAGAGAT 4500
4441 GGTAAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTATTAAGAGAT 4500
4501 ATGCAATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4560
4501 ATGCAATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4560
4501 ATGCAATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4560
4561 GACGAAATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4620
4561 GACGAAATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4620
4621 GAGACTGTAATAAATTTTGTGAGATCTAATTAAGAGATTAATTTGCTTAACTAT 4680
4621 GAGACTGTAATAAATTTTGTGAGATCTAATTAAGAGATTAATTTGCTTAACTAT 4680
4681 TATTCACAGATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4740
4681 TATTCACAGATTAATTAATTAATTTGCTTAACTATTAAGAGATTAATTTGCTTAACTAT 4740
4741 GATTTGCTTTTAAATTTGAGATCTTGAACATTAATTAATTTGCTTAACTATTAAG 4800
4741 GATTTGCTTTTAAATTTGAGATCTTGAACATTAATTAATTTGCTTAACTATTAAG 4800
4801 GAATATTAATTAATTAATTTGCTTAACTATTAAGATTAATTTGCTTAACTATTAAG 4860

Db	4801	GAATATTTTATTAATTAATGATGCTATTCGTACAGTTGCTTATTTAATCGAATAATTCCTATA	4860
Oy	4861	AATATATATGTTTATAGTCACAAGATTTGAGAAATTACCCCTTTAAAGTTAAAAAGAGTTT	4920
Db	4861	AATATATATGTTTATAGTCACAAGATTTGAGAAATTACCCCTTTAAAGTTAAAAAGAGTTT	4920
Oy	4921	AGTCATTTATTTTGATCGAAAGATTATTAAGAAGAGTTTAAATGTTTAAACAAATGTAT	4980
Db	4921	AGTCATTTATTTTGATCGAAAGATTATTAAGAAGAGTTTAAATGTTTAAACAAATGTAT	4980
Oy	4991	TCACACAGTTGTTTGATTAATGAGTTCCTGCAATTTAGAGTCTTATCGAAAAAGAAATA	5040
Db	4991	TCACACAGTTGTTTGATTAATGAGTTCCTGCAATTTAGAGTCTTATCGAAAAAGAAATA	5040
Oy	5041	CGTAGATATCCATTTATTAAGCCGAAAAGATATTTATCAAGAAAGCATTTAGTACGTTG	5100
Db	5041	CGTAGATATCCATTTATTAAGCCGAAAAGATATTTATCAAGAAAGCATTTAGTACGTTG	5100
Oy	5101	TATTTGATGAATAATTTGCGCTTAAACTATATGTATATTAATTAAGAAATTTCAAAAGCAG	5160
Db	5101	TATTTGATGAATAATTTGCGCTTAAACTATATGTATATTAATTAAGAAATTTCAAAAGCAG	5160
Oy	5161	TAGAGCTTAAAAATGCGATTAATTTAGTGTATTTGTCAGTTTATTAATGATTAATTAATTT	5220
Db	5161	TAGAGCTTAAAAATGCGATTAATTTAGTGTATTTGTCAGTTTATTAATGATTAATTAATTT	5220
Oy	5221	TAAAGTATGTTATGAAGCAATTTATCAAAATTAATTAATTAATTAATTAATTAATTAATGA	5280
Db	5221	TAAAGTATGTTATGAAGCAATTTATCAAAATTAATTAATTAATTAATTAATTAATTAATGA	5280
Oy	5281	TAGATGATGCGCTGTAGATGATTTCTGCTAAATATGCAAGAAATATGCAAAAAAGATA	5340
Db	5281	TAGATGATGCGCTGTAGATGATTTCTGCTAAATATGCAAGAAATATGCAAAAAAGATA	5340
Oy	5341	AAAGGTAAAAATTTTTTCACTAATCAATGAGATATCAATGCTAGAAATCATGGA	5400
Db	5341	AAAGGTAAAAATTTTTTCACTAATCAATGAGATATCAATGCTAGAAATCATGGA	5400
Oy	5401	TAAAGCGAGTACAGCTGCAATTAATTAATGTTTGAATCTGAGATGTTGTATATGTA	5460
Db	5401	TAAAGCGAGTACAGCTGCAATTAATTAATGTTTGAATCTGAGATGTTGTATATGTA	5460
Oy	5461	GATTAGTGAAGAAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5520
Db	5461	GATTAGTGAAGAAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5520
Oy	5521	TGTACGCTACTTTTTTCAGAAAAATTAATAATTTTGAAGTAATCAATATTTGATTT	5580
Db	5521	TGTACGCTACTTTTTTCAGAAAAATTAATAATTTTGAAGTAATCAATATTTGATTT	5580
Oy	5581	TTGGAAGCAATTAATTAATCCGTGCAAGCATGGAAGAAAAATTTTAATGAAATTTGATTAATA	5640
Db	5581	TTGGAAGCAATTAATTAATCCGTGCAAGCATGGAAGAAAAATTTTAATGAAATTTGATTAATA	5640
Oy	5641	ATTAATATTTTTCCTACTCCTGTTTGTAACATATTAAGAAAAAGATACATAACAGATCTTT	5700
Db	5641	ATTAATATTTTTCCTACTCCTGTTTGTAACATATTAAGAAAAAGATACATAACAGATCTTT	5700
Oy	5701	TTCAAGAGAAATCAATGTTTAGAGAGAAATTTACTTTTAAATCTGATTTATTTAAAGATA	5760
Db	5701	TTCAAGAGAAATCAATGTTTAGAGAGAAATTTACTTTTAAATCTGATTTATTTAAAGATA	5760
Oy	5761	TAGATTAAGATTAATTTTGAATGCAACATTTTATTTTAAGAGAGATATCTTAAGTA	5820
Db	5761	TAGATTAAGATTAATTTTGAATGCAACATTTTATTTTAAGAGAGATATCTTAAGTA	5820
Oy	5821	CAGTAAATTTCTTTTAAAGAGGTGTTTGTGCAATTTGAAAAATTTGCAAAAAACAAGTA	5880
Db	5821	CAGTAAATTTCTTTTAAAGAGGTGTTTGTGCAATTTGAAAAATTTGCAAAAAACAAGTA	5880
Oy	5881	TAGATTTGTTTAAAGCAAAATTAATGCTGAGATTTTGAATGATCAATTTGTTAAAGATCTA	5940

Db	5881	TAGTATTTGTTAAGCAAAATATATGCTGAGATTTTGACGTAATCAATTTGTTAAAGATAC	5948
Qy	5941	TACGTTGGCAAGTATTTATTTATATAGCTTACTTAATGTTTAAATACGAAAACAGCTATTT	6000
Db	5941	TACGTTGGCAAGTATTTATTTATATAGCTTACTTAATGTTTAAATACGAAAACAGCTATTT	6000
Qy	6001	TTGACAAATTTTATTTTATTTTGAATCTCTTAATAAAAAATATATTTTAACTGTGTTAAAG	6066
Db	6001	TTGACAAATTTTATTTTATTTTGAATCTCTTAATAAAAAATATATTTTAACTGTGTTAAAG	6066
Qy	6061	TATCTAACAAAAATTCCTTGCTAATAAAATTTTGTATAGAAATGTTGCGAACAAAGTTT	6120
Db	6061	TATCTAACAAAAATTCCTTGCTAATAAAATTTTGTATAGAAATGTTGCGAACAAAGTTT	6120
Qy	6121	TTAAAAAAATATATATGTTATATATATAGGAAGATATCATGATACTATTAAGTAAAAATTTCTA	6180
Db	6121	TTAAAAAAATATATATGTTATATATATAGGAAGATATCATGATACTATTAAGTAAAAATTTCTA	6180
Qy	6181	TAAATTTGACCTATATATATATATAGAAAAATTTTATCTAAATGTATAGATAGCATTTGTA	6240
Db	6181	TAAATTTGACCTATATATATATATAGAAAAATTTTATCTAAATGTATAGATAGCATTTGTA	6240
Qy	6241	ATCAGACCTAACAAACATATAGAAATTCCTCGGGAATGACGTATACGGATTAATTCGG	6300
Db	6241	ATCAGACCTAACAAACATATAGAAATTCCTCGGGAATGACGTATACGGATTAATTCGG	6300
Qy	6301	AAGAAATTTGTTTATGACATATGCGAAGAAAGATAGTCGATTCGTTATTTTAAAAAAGA	6360
Db	6301	AAGAAATTTGTTTATGACATATGCGAAGAAAGATAGTCGATTCGTTATTTTAAAAAAGA	6360
Qy	6361	ACGGGGGGCTATCAGATGCCCCGTAAATATGAGCATTAAGTCGGCCAGAGGTATCTACTTG	6420
Db	6361	ACGGGGGGCTATCAGATGCCCCGTAAATATGAGCATTAAGTCGGCCAGAGGTATCTACTTG	6420
Qy	6421	CTTTTATATGACTCAGATGATTTTATTTTCAATTCGGAGTTCAATCCAGCGTTTACAGAGGAA	6480
Db	6421	CTTTTATATGACTCAGATGATTTTATTTTCAATTCGGAGTTCAATCCAGCGTTTACAGAGGAA	6480
Qy	6481	TTGAGAGAGAAATGCCCCCTTGTGACAGTGTGCTTATATGATAGGGTATGATCTTCGGGGC	6540
Db	6481	TTGAGAGAGAAATGCCCCCTTGTGACAGTGTGCTTATATGATAGGGTATGATCTTCGGGGC	6540
Qy	6541	ATTTCCTTAACGACAGAGCCGCTTCCTAACAAATCAGGCGTTCCTGACGGCAGGAATGTTT	6600
Db	6541	ATTTCCTTAACGACAGAGCCGCTTCCTAACAAATCAGGCGTTCCTGACGGCAGGAATGTTT	6600
Qy	6601	GTAAGAAAGCTCTAGAGGCGGATGCTCATGCTTGTGTGCGCCCTGTAAATTAATCTCTATA	6660
Db	6601	GTAAGAAAGCTCTAGAGGCGGATGCTCATGCTTGTGTGCGCCCTGTAAATTAATCTCTATA	6660
Qy	6661	AAAAAGAACTATTTGAAGATTTTGCATTTTGAAAAAGGTTAAGATCTCATGATGATGATACT	6720
Db	6661	AAAAAGAACTATTTGAAGATTTTGCATTTTGAAAAAGGTTAAGATCTCATGATGATGATACT	6720
Qy	6721	TCACCTTATCGCTTGCTCTATGATGTTAGAAAAAGTTGCAATAGTTAAGAGTGTCTGTACT	6780
Db	6721	TCACCTTATCGCTTGCTCTATGATGTTAGAAAAAGTTGCAATAGTTAAGAGTGTCTGTACT	6780
Qy	6781	ATTATGTTGACCCGGAATAATAGTATCACACTTTAGCATGATGACCATTCGCTTCATTT	6840
Db	6781	ATTATGTTGACCCGGAATAATAGTATCACACTTTAGCATGATGACCATTCGCTTCATTT	6840
Qy	6841	GCCCTACTGGAATTTCAAAAATGAAATGCAATCGACTTCTATGAAAGTATAGAGATTAAGAC	6900
Db	6841	GCCCTACTGGAATTTCAAAAATGAAATGCAATCGACTTCTATGAAAGTATAGAGATTAAGAC	6900
Qy	6901	TCTTACTAGAGTGTATTCGTTCAATTTTAAAGCCTTGCTGTGTTTGTTTTAAAGCAATATA	6960
Db	6901	TCTTACTAGAGTGTATTCGTTCAATTTTAAAGCCTTGCTGTGTTTGTTTTAAAGCAATATA	6960
Qy	6961	ATCATTTGTTAGCAAAACAGCAAAAAGAGCTT 6992	
Db	6961	ATCATTTGTTAGCAAAACAGCAAAAAGAGCTT 6992	

	RESULT 2	
LOCUS	Afl18389	
DEFINITION	Afl18389 15401 bp DNA linear BCT 11-APR-1999 Streptococcus suis Cps2A (cps2A), Cps2B (cps2B), Cps2C (cps2C), Cps2D (cps2D), Cps2E (cps2E), Cps2F (cps2F), Cps2G (cps2G), Cps2H (cps2H), Cps2I (cps2I), and Cps2U (cps2U) genes; complete cds; Cps2K (cps2K) gene, partial cds; and unknown genes.	
ACCESSION	Afl18389	
VERSION	Afl18389.1 GI:4580620	
KEYWORDS	.	
SOURCE ORGANISM	Streptococcus suis Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus. . (bases 1 to 15401) Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A. Identification and characterization of the cps locus of Streptococcus suis serotype 2; the capsule protects against phagocytosis and is an important virulence factor <i>Infect Immun.</i> 67 (4), 1750-1756 (1999)	
JOURNAL PUBMED REFERENCE AUTHORS TITLE	2 (bases 1 to 15401) Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A. Direct Submission Submitted (07-JAN-1999) Department of Bacteriology, DLO-Institute for Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The Netherlands	
FEATURES source	Location/Qualifiers 1..15401 /organism="Streptococcus suis" /mol_type="genomic DNA" /strain="10" /serotype="2" /db_xref="taxon:1307" <1..721 /note="Orf22" codon_start=2 transl_table=1 product="unknown" protein_id="AAD2444_1" db_xref="GI:4580621" translation="SLIDHMMEVWEASKSAGSCSPQAYQAFFGCANIYVTIT GGLSGSFNAARVARDMYIEHPNVNIHLISLSASGEEDLVHQINKLISGDLPPOV VEALITHREHSKLFLVALKNVLVNKGRLSTLGTVGVGLINIENVGAASEGKEILLQ KARGKSVTALEFEEMKRGXDGRIVMAHHNNAKFGQSSELVKASFPAVIDEVAT IGLCFYVAEGGLMGYEVRKA"	
CDS	complement(820..2079) note="Orf21" codon_start=1 transl_table=1 product="unknown" protein_id="AAD2444_1" db_xref="GI:4580622" translation="WKKTQVIILQDLTGTIEHRFRGGKPLSIROLREQHCSDTYQ KMELRYQNKIYAVERGSYYILEDPDFQHTRCAQSYRSIRITYIEDPRICLKESLLG RNLYLFNYHQOEGIAELISSVSLLMDHYHTKKDYLTAVSQQLYLTTOMETLA GETELLLENPTYSRMIELRHOGIPYCOTIERLDLIDELESIFORTKIFPFTIPRA LNHGPGSTDYDIATRTAIYAAKOVDVYIIEDDYADPPSSSHSLPHYLDGNRIYATQ SETPTEPALRRIGAISIPLNRADIPIKHKSLIDYPNTLMOKALSIVINDMFPARNQQ HLHHIYHAQMNKIKDCLETXYLANIPYRPKGSVTFQLSKGISPSIQMFCKCYFFSG QRADFQIPEFDKFADKLEQFVRYINE" 2202..2936 note="Orf2X" codon_start=1 transl_table=1 product="unknown" protein_id="AAD2446_1" db_xref="GI:4580623" translation="MKIILPNKAEVTNLNENASYLLSDRSKPVLDAISOFDVKMMAP	
CDS		

gene
CDS
3041. .4486
/gene="cpsa2a"
3041. .4486
/gene="cpsa2a"
/function="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="Cpsa2a"
/protein_id="AAD24447.1"
/db_xref="GI:4580624"
/translation="MKRRSGRSKSKFKLVNPAALIGVYSTLCTPLVPMVRYNIIIFR
LYNTIVLLVAGVALGLMLMKRAKIFPALLVSVLTVTSVGIYGMQVYKFSRL
NSNSTPSEYEKSLIPANSDITDVRQLTSLIAEYDQDITALLDDISRMESTQATL
KSGSYSLTAQSMNGESQAMVFNQVFTNILENEDGFSGKKIYFSFVTOVEIAT
KQVSDFNITVIGSIDAYGPISITVSRSDVNIIMTVNRAITHKILITTPDPDSYVAFADG
QONQYDKLTGAGIYGVNASVHTLENPQDITISYVNLFIPTFQILIDVGIDVYNDQ
EFTLHGHNYHPVQGVHLNDSQALGFRERYSLTGGDNRGKQKEVITALLIKMSTP
ENLNKYAALISHEGSIQTDSLTTMSLVNTOLESGTQPTVESQALTGTGRBDLSY
AMPQSOLYMEINDSDSQKAAIQSVLYEK"
4504. .5193
/gene="cpsa2b"
4504. .5193
/gene="cpsa2b"
/function="putative role in chain length determination"
/codon_start=1
/transl_table=11
/product="Cpsa2b"
/protein_id="AAD24448.1"
/db_xref="GI:4580625"
/translation="MNNOEVNAIEIDVLFLLKTIWRKKFILLTAVITAGLAVYSF
LVTPOYSTTRIIYVSONVEAGAGLTNQELQAGTYLAKDREIILISQDVLTVATELN
LKESPEKTSISIVIPDTRIVASIVRADPPEARAIENSLTFVQKVEATKYSDVTT
LEAVPAEPPTPMTKRNILLGLAGIATGTVLWVEIVLDVRKPODIEEVMGLTL
LGIVPDSKKTK"
5203. .5880
/gene="cpsa2c"
5203. .5880
/gene="cpsa2c"
/function="putative role in chain length
determination/export"
/codon_start=1
/transl_table=11
/product="Cpsa2c"
/protein_id="AAD24449.1"
/db_xref="GI:4580626"
/translation="MAMLEIARTBEGVNTKEEYFNAIRTNIOLSGADIKVNGITSVK
NBSGTSITTAASIALIYARSGYKTVLVDADIRNSVMPGFPYIKTIGLIDYLAGTDL
SGLGCDTIDIPNITVIESGKVSNPFTALLQSKNFNLATIRRYDYIVTDCPPLGLVY
DAALIAQCDMAVAVEAGNVKCSSLKVKYEQLEQTGPFLGVLINKYDILATEKYSBY
GNVYKKA"
5919. .6650
/gene="cpsa2d"
5919. .6650
/gene="cpsa2d"
/transl_table=1
/codon_start=1
/product="Cpsa2d"
/protein_id="AAD24450.1"
/db_xref="GI:4580627"
/translation="MIDIKSHVIFGVDDGPPTIEESLSLISEAVRQGVRYVATSHRR
KMPFEPEKIMINIFOLKEAVAEVPEIRLCYGAELIYSSKDIISKEKKVPLINSS
CYLILFSDTPPWKEIOEAVNEMTLGLTVPVLAHERYDALAFQSEVEKXIDGCYTVT
QVNSHNRALKPAIGBRAGEKKRTYRFLFDGLVHCVASDMNLYSRBPFRBEAYQLVAK
KEYGEARAKALFKKAPLILIKQVQ"
6675. .8054
/gene="cpsa2e"
6675. .8054
/gene="cpsa2e"
/note="putative glycosyltransferase"

Db	13703	TCCTTCAGATTCCATCAACCGGATATATGTTGGAAATACGCAAGACCAAGATGATGAATTTAAA	13766
Qy	5352	ATTTTTCCTACCTATCATAGTGGAGATCAAAATGCTAGAAATCATGGAATTAAGCGAGT	5411
Db	13763	CTTTTCGGTTCACCAATATGGTGTGTTTCAMAGCAAGAAATTAACGATATCAAAAATATAC	13822
Qy	5412	ACAGCTGAATATATTTATGTTGTGTCCTGATGACGTTGGTATGTGATATAGTAGAA	5471
Db	13823	ACAGCAATATATATATGTTGTAGATTCGAGATATATGTGACCGCACATTTGTTGAG	13888
Qy	5472	AAATATATTTTATATATTTAAAAAGTAGAAGTATTTATCTGTTGTTGTACGCTACT	5531
Db	13883	TCCTTATACGCTGTTTAAAAAGAAATGATAGTATTTGTGCGGAAGGTTATCTTGCTACT	13947
Qy	5532	TTTTCAAAAAATATATAATATTTTGAAGTGAATATCCAAATATGATTTTGAGCAAT	5591
Db	13943	TTTGATGGAATATATATCAAGATCTGACGCAAAAGTGTCAATATGATTTGGAAGAAAT	14002
Qy	5592	AATATCGTGAAGACATGCGAAGAAAAATTTATATTTGATATATATAATTTTT	5651
Db	14003	AAAGAGGTGGACCTTAGAAATGAAATTTTCCAAATCATATATATGACGGTATCTTT	14066
Qy	5652	TCTACTCTGTGTTGTAACCTATATTAAGAAAGATACATACAGATCTTTTCAAGAAAT	5711
Db	14063	AATATGCCCTTGTCCAACTTATATAGAAATATATATATAACCAAGCTTTTGACCTGA	14122
Qy	5712	CAATGTTAGAGAAATTTACTTTTAAATCTGCATATTTAAGAAATATAGATGATT	5771
Db	14123	CAGTGTTAGAGAGACCTTATATATTTAATCTAAATTTAAGAAATATAAAAAAGTC	14182
Qy	5772	AGTATTTGACTGAACATCTTATTTTATAGAGAGATATCTAAGTACAGTAATCT	5831
Db	14183	CGCTATGTTAACAGAAATCTTATTTTGCAGAGAAAGTTTCAAAATGATCAAAATAG	14242
Qy	5832	TTTAAAGAGGTGTGTTTTCGAATGGAATTTGCAAAAACAAGTATAGATGTTT	5891
Db	14243	TTTAAATATGATGTTTATTTATCAATTAGAAAATTTGAAAGAAAAAATTGATTTGTTT	14302
Qy	5892	AAGCAATATATGATGAGGATTTTGAAGTATCAATTTGTTAAAGATATCTATAGTGGCA	5951
Db	14303	GTTAAATATTTGCTGGGACATATGAAATTTTCGTTTAAAGACGCTACAGTGGCAT	14362
Qy	5952	GATTTTATATATGCTTACTAATGTTTAAATACGAAAACAGTCTATTTTTCACAAATTT	6011
Db	14363	ATATTTATATATGCTTATTTAAAGTTCAAAAATGAGAGAAATCGCTTCAAAAGAAATG	14422
Qy	6012	TTAATTTTAAAGAAATCTTATATAAAAAATATATTTAACTGTGTAAAGATATCAACA	6071
Db	14423	CATATATTT--AGTATTTTATACAAATAGGCATTTTATGATCTCTAAGTATTTAAACGA	14488
Qy	6072	AATTCCTTGTCTAAAAATTTTGTATAGAAATGTTTCGACAAAGTTTAAAAAAATA	6131
Db	14481	CGTCTCTGTTTAAAAAGAAATATGTAATATATGTCATATATTTGTTTAAAAATTT	14548
Qy	6132	TATGCTTATATAGGAAGATATCATGATATCTATTTAGTAAAAATTTCTATTAATTTGACT	6191
Db	14541	TTTTAATATCTTTAATTTAGGAGAAAAAAATATATCATTTTCTATCATCTGCCCA	14600
Qy	6192	ATATATATATGTAAGAAATATTTATCTTAATATGATATAGCATTTGTAATCAGACCTAC	6251
Db	14601	ATTTACAAATGTTGAACAATATCTATCTCAAGTGTATATTAATATGATATCAACCTAC	14660
Qy	6252	AAACATATAGAGATCTTCTGCTGATGACGGTAGTACGATATTTCCGAAAGAAATTTGT	6311
Db	14661	AAACATATATAGATCTTCTGCTGATGACGGTAGTACGATATTTCCGAAAGAAATTTGT	14720
Qy	6312	TTTGCATATGCGAAGAAAGATATGTGCGATTTCTGTTATTTAAAAAGAAACGGCGGCTA	6371
Db	14721	TTTGCATATATGCGAAGAAAGATATGTGCGATTTCTGTTATTTAAAAAGAAACGGCGGCTA	14780
Qy	6372	TCAGATGCCCGTAATTATGSCATATAGTGCAGCAAGGAGCTACCTAGCTTTTATAGAC	6431
Db	14781	TCAGATGCCCGTAATTATATGSCATATAGTGCAGCAAGGAGCTACCTAGCTTTTATAGAC	14840

QY 6432 TCAGATGATTTTATTCATTGAGATTGATCCAAAGTTTACAGAAAGCAATTGAGAGAG 6491
| | | | |
Db 14841 TCAGATGATTTTATTCATTGAGATTGATCCAAAGTTTACAGAAAGCAATTGAGAGAG 14900
| | | | |
QY 6492 AATGCCCTTGTGGCACTGCTGCTGTTATGATAGAGTATGATGCTTCGGGGCATTTCTTAACA 6551
| | | | |
Db 14901 AATGCCCTTGTGGCACTGCTGCTGTTATGATAGAGTATGATGCTTCGGGGCATTTCTTAACA 14960
| | | | |
QY 6552 GAGAGAGCCCTTCCTCAATATGAGGCTGTTCTGAGAGGCGAGAAATGTTGTAAAAAGCTG 6611
| | | | |
Db 14961 GAGAGAGCCCTTCCTCAATATGAGGCTGTTCTGAGAGGCGAGAAATGTTGTAAAAAGCTG 15020
| | | | |
QY 6612 CTAGAGAGCGAGTATGATGCTTGTGTGCTGCTGTTATTAATCTTATTAATAAGAACTA 6671
| | | | |
Db 15021 CTAGAGAGCGAGTATGATGCTTGTGTGCTGCTGTTATTAATCTTATTAATAAGAACTA 15080
| | | | |
QY 6672 TTTGAAGATTTTGCATTTGAAAAAGGTAAAGATTCATGAAGATGAATTAATCTTATTCG 6731
| | | | |
Db 15081 TTTGAAGATTTTGCATTTGAAAAAGGTAAAGATTCATGAAGATGAATTAATCTTATTCG 15140
| | | | |
QY 6732 TTGCTCTATGAGTTAGAAAAAGTTGAATAGTTAAAGAGTCTGTACTATTATGTTGAC 6791
| | | | |
Db 15141 TTGCTCTATGAGTTAGAAAAAGTTGAATAGTTAAAGAGTCTGTACTATTATGTTGAC 15200
| | | | |
QY 6792 CGAGAAATATGATCACTCAACTTCTAGCATGACTGACCATGCTTCATTCGCTACTGGA 6851
| | | | |
Db 15201 CGAGAAATATGATCACTCAACTTCTAGCATGACTGACCATGCTTCATTCGCTACTGGA 15260
| | | | |
QY 6852 TTTCAAAATGAAAGCAATGCACTTCTATGAAGATGAAGAGATTAAGCTTACTAGAG 6911
| | | | |
Db 15261 TTTCAAAATGAAAGCAATGCACTTCTATGAAGATGAAGAGATTAAGCTTACTAGAG 15320
| | | | |
QY 6912 TGTATTCGTTCACTTTTATGAGCTTGTGCTTTGTTTATGCAAAATATATCATTTGTTG 6971
| | | | |
Db 15321 TGTATTCGTTCACTTTTATGAGCTTGTGCTTTGTTTATGCAAAATATATCATTTGTTG 15380
| | | | |
QY 6972 AGCAAAACAGCAAAAGAGCTT 6992
| | | | |
Db 15381 AGCAAAACAGCAAAAGAGCTT 15401
| | | | |
RESULT 3
AY057915 17468 bp DNA linear BCT 01-NOV-2002
LOCUS Streptococcus thermophilus eps type III operon, partial sequence.
DEFINITION
ACCESSION AY057915
VERSION AY057915.1 GI:24473733
KEYWORDS
SOURCE Streptococcus thermophilus
ORGANISM Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS 1 (bases 1 to 17468)
TITLE Rallu, F., Ehrlich, D.S. and Renault, P.
JOURNAL Diversity of eps operons in Streptococcus thermophilus
REFERENCE 2 (bases 1 to 17468)
AUTHORS Rallu, F., Ehrlich, D.S. and Renault, P.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Genetique Microbienne, INRA, Domaine de
Valvert, Jouy en Josas 78352, France
FEATURES
source
1..17468
/organism="Streptococcus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:1308"
<1..17075
/gene="eps type III operon"
<1..492
/gene="decd"
<1..492
/gene="decd"
CDS
/gene="decd"

gene
CDS
/codon_start=1
/transl_table=11
/product="Deod"
/protein_id="AAL23725.1"
/db_xref="GI:24473734"
/translation="RELIVDYGVKULIRVGTAGSLNEDVHYRELVLQAQAAATNSNIIR
NDMPQYDPPQIASFRLDKAHHIAKRLPMTTHVGNVLSDDVPYSVPEKNIELGKGV
KAVEHEAALYYLAAQHVDAALATWTTISDSLVNPEBDTTAEERONTFTDMKVGLETL
IAD"
828..2288
/gene="eps3A"
828..2288
/gene="eps3A"
/codon_start=1
/transl_table=11
/product="Eps3A"
/protein_id="AAL23726.1"
/db_xref="GI:24473735"
/translation="MSRRNRRKORTGNRSMGVNVGLTLYALALVLLFTMPVNF
LSFRPLNITITGLVLALISFLQRTKSPLYTTVTVVTSVLSVLCIGPKOMIDI
TNRNQTAFFSEVMSIVPKESDIDKDVSLTSVQAPTKVDKNNIEILMSALKOKKV
DKVDVVASVQEAADNLSGKSKAMVLSGSVASLLESVDSPASNLKTIYYKIKKGN
NNSANQVDSKVENIYISGIDYGSISTVSRSDVNIIMTVNMNTHKILITTPRDAVVK
IPGGANQVDKLTHTAGIYGVETSEOTLEDLYGTIKIDYARINFTSLKILIDLGCVTV
HNDQAFSLHGKRPVPGDIGNSRDAGFVPERYSIDGSDNRKDKOEVSAIVNK
HSLKSVSIFTSIVNNLQDSVQTNISLDTINALATPQDLSGSKFTVTSQAATGTSSTG
QLTSYAMPNSLSLYMKLNDNSVASQATKNLMEEK"
2289..3020
/gene="eps3B"
2289..3020
/gene="eps3B"
/codon_start=1
/transl_table=11
/product="Eps3B"
/protein_id="AAL23727.1"
/db_xref="GI:24473736"
/translation="MIDVHSHTVPPVDGPKTLSESLDIGSVAGVYRKIVSTSHR
KMFETPEBDKTRANFSKVAKABALYPDTITTYGSELVTTDITVEGLKLNLRPMNT
QFALIEFSARTSMKEIHSGLSNVLRAQVPIVAHIERVDALENADRVETINMGCTT
QVNSSHVLRPKLFGDKKVRKRRVFFLEKNLVHVVASDMHNLGPRPPMKDAYELVK
KNVGSRAKNLFTENPKTLLENQYL"
3029..3721
/gene="eps3C"
3029..3721
/gene="eps3C"
/codon_start=1
/transl_table=11
/product="Eps3C"
/protein_id="AAL23728.1"
/db_xref="GI:24473737"
/translation="MNQDWTKSDEIDVALLHLKMTYKLLIFTAFFPAFBSFLGTYF
FIQPTYSTTRIVYVQAQDNNNLSAQDQAGTYLNDYKEIITSDNIVASEVTKVDK
NISEALSKWVSVINIPTDTRLISIVNAQTGDAQCLANKRVHVASAKKIKKTVKXVDV
TWLEBAKLPESPSPIKLVNLGAVLGGFLAVGVLVLEILLDDRVRRBEDVDALGM
ALLGIYPTDKIT"
3731..4489
/gene="eps3D"
3731..4489
/gene="eps3D"
/codon_start=1
/transl_table=11
/product="Eps3D"
/protein_id="AAL23729.1"
/db_xref="GI:24473738"
/translation="MPLKLKVSKNVFAKQTEYYNAIRTNIOFSGAQIKVIAISSE
AGEGKSTISNLAISFASVGLRTLLIDATRNSVRFGSTKSNPYYVLIIDTPVGVLI
NETTCQTDISGLDVIVASGVPVNPFTSLQNDNRHRLMEVARSYDVYLIIDTPVGVLI
DAVIAHQDASLIVTAAGIKRRPTKAVEQLEGSQPLGVLANKVDWTVKXGYSY
GSYSGYSYGSYGEYRKTDQTEBGRRAHRRRG"
4546..5913
/gene="eps3E"
4546..5913
/gene="eps3E"

```
/note="undecaprenylphosphate glucosyltransferase"
/codon_start=1
/translation_table=11
/product="Eps3F"
/protein_id="AAL23730.1"
/db_xref="GI:24473739"
/translation="MKKEKQIRRIEIGIIQIVVIFPAMVASKIPYTEITQGSIVLIG
VHVSVYISYYENLKRYGDELATVYCFIPALIFATLSPFADGSPSISRGIL
VTMTISGVLLVYTNVTLKYPRSSITRRKSNKILLISQARLDNVLSPKMDMDRI
SAVCUDNPIFTDPTIKSVKPEMLIEATHSVVDVLIINPSQYKIMDYASFPELML
IPVSIINLALFESGSKHIOQLGPRKVVFTSFYSYDILAKRFLDICGALVGLV
CGIVGIFLYPLIRKDGPAIPQYVGENGRIFPKYFRSMRYDAEIKKIMLAQNM
SGCFMMDNDPRTIKGIFIRKTSIDELPQFMNVLKGDMSLVGTDPPTVDEYKTYPE
OKRSLSPKGIITGLMOVSGRSEITDDEVVKLVAVIIDGWTIWRDIQILKTKIVVM
KQAK"
gene
5946..6264
/gene="orf3A"
/pseudo
5946..6264
/gene="orf3A"
/note="orf3A, putative rhamnosyltransferase"
/pseudo
/codon_start=1
/translation_table=11
6275..7647
/gene="eps3F"
/pseudo
6275..7647
/gene="eps3F"
/note="Eps3F, putative undecaprenylphosphate
glycosyltransferase"
/pseudo
/codon_start=1
/translation_table=11
7682..8131
/gene="eps3G"
7682..8131
/gene="eps3G"
/note="eps3G"
/note="putative galactosyltransferase enhancer"
/codon_start=1
/translation_table=11
/product="Eps3G"
/protein_id="AAL23731.1"
/db_xref="GI:24473740"
/translation="MKICLVGSSGSHLAHLMKPKWQBERFWTTPDXEDARSILKD
EKMYSCHFPITNRLKNLKNITFLAIVLKKEKPDVLIISGAAPVFFYIGKLPGAKT
VIEVFDRIIDKPVTKLVYPVDKFIQVEMKQVYPRAINLSIF"
8131..8637
/gene="eps3H"
8131..8637
/gene="eps3H"
/note="eps3H"
/note="putative galactosyltransferase"
/codon_start=1
/translation_table=11
/product="Eps3H"
/protein_id="AAL23732.1"
/db_xref="GI:24473741"
/translation="MIFVTVGTHQOPNNRIQEVDLVETGVKEEVFIOTGYSIYEP
KFCOMRLISFDQMSKFMQKADIIITHGSPATPMSAITNGKPIVPRQEKFEHND
HOYDFARNVAKRMGTIILVEDVDKIGETIENNSIVABLEYDMGNKKCFMNDLEGI
VYLDLNVK"
gene
8642..9358
/gene="eps3I"
8642..9358
/gene="eps3I"
/codon_start=1
/translation_table=11
/product="Eps3I"
/protein_id="AAL23733.1"
/db_xref="GI:24473742"
/translation="MNYAKTCFLIIAHNNMGQIKLIECLDSQTHDIFVHVDKSDP
DSYFIDSVTKSLKFEFTVWDGSPQOVEMFLIEQVYIEGYVYHIIISGADJPL
KNKEIDLPFEENKGEFIIYDNNLTEDNLISRTKYHYHFLQNYRRRYTEWNNKNSFF
```

```
TCERVSVLVQIIIFGVNRVKNLDWQIKYGSQWVSITNELVKTILLEHEKITSIFYOTK
CKNHRDSIFISFNSQVSN"
gene
9374..10474
/gene="eps3J"
9374..10474
/gene="eps3J"
/codon_start=1
/translation_table=11
/product="Eps3J"
/protein_id="AAL23734.1"
/db_xref="GI:24473743"
/translation="MKILQIPGGLFSDGINSFIYVYVYAMDKSENDIRVLA TNABE
STLQRYKEGCEVQSIPIYRKQNI IKYFELFRYIAEKIDIVVHSSSAIMSVELAA
KLAGCKVRJAHSHNITCENORVDQILRPEDKLYTEFCGSGAGKMLPGRNRFVYIP
NRSILKQVRYNOKRLOVRKHIIPDDALVIGVGRFNOKKHVLIQI FYEFHKYIP
KSYVLIGTGDVTYRKQOVTELEENYVIFIGAVNNVDYLSAFPVMLPSIYELP
SVVIEQWISGLPCLISDNTNECKTSLVEPESITETTPWARDIVELKIEDRNIND
KIFKEVRDAGYDIDBDKKLKEIYESLYSFIK"
10496..11166
/gene="eps3K"
/pseudo
10496..11166
/gene="eps3K"
/note="Eps3K, putative glucosyltransferase"
/pseudo
CDS
gene
Query Match 13.7%; Score 955.6; DB 1; Length 17468;
Best Local Similarity 65.8%; Pred. No. 1,36-87;
Matches 1484; Conservative 0; Mismatches 729; Indels 41; Gaps 5;
QY 1 ATGCCAAGAAATGGCATTTATGATATGATAGCAGTGCATTTCTGCATCTTAA 60
DB 6294 ACCGCAATTTAGACTGCGGCTTTAGAAATGATGGGTAGTAGCGCTGTCCATTA 6553
QY 61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGAAATTTTATCATTAATGATGTT 120
DB 6354 CCACCACTTGCACCTCAAAAGATGATCAGTAGGACGATTAATGATTAATGCCCTGC 6413
QY 121 ATTATTTTGCATTTTATATATCTCGATGCGCACTTAATTTAGATATAGGTAATCTGA 180
DB 6414 ATTTGTGCTTTTATTTCTACGATGTCATCAAAATTTGAGATAGGCTACTCTTAA 6473
QY 181 TGAGTTTGAATAAAACATTTACTATAGTATAATTTTGAATTTTCTTACGCGAGTAT 240
DB 6474 TAGAGTTTGAACGGTCCGCTAGATATTTTGAATTTTCAATTTCTTCAATTTT 6533
QY 241 CATTTTGTGAGATAATTTTGCACCTTCAAGAGTGTGCGGTATTTTGCATTTA 300
DB 6534 CTTTTTATTTGGAGATAGTTTATCATTAACGCGAGTGTGCGGTATTTTACCATTC 6593
QY 301 TAACTTCGTTTGTGATACCTATTTAAGGTAATTTAAGCAGTTTAAGATAGCTTTC 360
DB 6594 TTACCTTTATCATCGTTTATATATATCACTGCATCTGTGTGATATCAAAATACCTCTTCT 6653
QY 361 TATTTTGCACATCTATCAAAAAAAGACGATCTTATTAACAGCGTCAAGACATGGGAAA 420
DB 6654 TTATGCTGCTGAACAAGAAATATCTTATGCTATTAACATCTAGAGATTCGAT 6713
QY 421 ATATGCAATTTTATTTGATACATATAACAATTTCAAAAAATCTTGTGCAATTTAG 480
DB 6714 TATGAGAGGGGCTTTGATAGGATCAACCTTCTACAAATATTTGGGGGATATGTC 6773
QY 481 TTTTAGGTACAGAAATAGATAAATTAATTTATCATTTACCGCTCTATTTATCTGTGAAG 540
DB 6774 TCAATGG---AGACGGTATGATGCGCTTTCCAGAGGAGATTCAGTAGTTCCTTTGATTT 6830
QY 541 AAGCTATAGATTTTCAACAAGGAGTGTGACACGCTTTATTAATTTACCAAGTG 600
DB 6831 ATGCAATAGATTTTGCAGCCATGAGATGTGTGACCATGTTTATCAACTTACGAGTG 6890
QY 601 AGTTTATGACGTAAGCAATTCGTTTCAAGTTTGAATTTGATTTAGTATGATGAAGG 660
DB 6891 AACATTACATCTCAAAACATCTTGTTCGATTTTGAAGTATGAGTATGATGATGAGTG 6950
```


QY 121 ATATATTTGCAATTTTATATCTGTATGCGAGTTGATTTGAGTATAGAGTATCTGA 180
DB ATTTTGGGCTTTTATTTCTCAAGTATGTCAATGAAATTTGAGATGAGGCTACTTAA 6473
QY 181 TAGAGTTGAAAAAACAATTATCTATAGATATATTTTGGCAATTTTCTTACGGAGTAT 240
DB TAGAGTTGAAACGGGTGCTAGATATGTTTGAATTTTGAATTTCTCTCAATTTCAATTT 6533
QY 241 CATTTTGTGGAGAAATATTTGCACTTTCAACAGTGTGCGGTGATTTTCCATTA 300
DB CTTTATTTATGGAGATATTTTATCATTCGACAGTGTGCGGTGATTTTATTAACATTC 6593
QY 301 TAAACTCTGTTTGTATACCTATTTTAACTGATTTATTAAGCAGTTTAAAGATCTTTC 360
DB TTACTTTTATCATCTTTATATTTATCACTGCACTGTTCTGATATCAAAATCTCTCTCT 6594
QY 361 TATTTTGCACATCTATCAAAAAAAGACGATTTCTATTAACAACGGCTGAACGATGGAAA 420
DB TTATGACTGCTGAACACAGAAATATCTTATGATTTACCACTACGATGAGATGGAT 6713
QY 421 ATATGCAAGTTTATTTGATCAGATTAACAATTTCAAAAAATCTTGTTCATGATG 480
DB TAAATGGAGGCTCTTTGAAATCGGATCACTTCTACCAATATTTGCGGTATTTGATG 6714
QY 481 TTTTATGATACAGAAATAGATTAATTAATTTATCATTAACGCTATATTTCTGTGAG 540
DB TCATTTG--AGACGATGATGTGGCTTTCCAGAGGAGTTCCAGTATCTTCTTTGATTT 6830
QY 541 AAGCTATAGATTTTCAACAAGGAGTGTGACACAGTCTTTATTAATCTACCAAGT 600
DB ATGCAATAGATTTGCAACCATGAAATGTGACACAGTGTGATCACTTACCGAGT 6831
QY 601 AGTTTATGACGTAAAGCAATCTGTTCAATTTTGAATTTGATTTGATTTGATTTGAT 660
DB AACATTCAGATCTCAAAATCTGTTTCCGATTTTGAAGTATGAGTATGATGATG 6891
QY 661 TTGATTAATTAATCTGTTTATCTGTTTACGCTGTAACCAAAAAATCACTGATAGT 720
DB TAGATTAATTAACATCTGATTTTATGAGCTTTTAAACAAATTAATAAATCAAGTTGAG 7010
QY 721 ACCATAGCATTTGATCTTTTCCAAATTTTATTAAGCTATGATATGATGAAC 780
DB ACCATAGCATCTGATCTTTTATCTCAATTTATCTCAAAATAGCATATCTTTCTTAAGC 7011
QY 781 GACTTTGATATATCTGAGCGGTAGTCTGGTTATTTATTTGTGTATATGTTCTATAT 840
DB GCATGTTGATATCTTTGGGCTTTGATTTGTTCTTAATTTGTGTGTGCTGCTGAGATTTG 7071
QY 841 TGTATGTTCCAAATTTATCTGATAGATATGTTGACCGGTATTTTGTCTCAAAAAGATTTG 900
DB TCTTATGCTCCATCTATTTGGAAGATGAGGACCGGATTTTGTTCAAAAACGATAG 7131
QY 901 GACAGATGAGCAGATATTTATCTTACAAATTTTATGATGATGATGATGATGATGATG 960
DB GGAAGATGAGCAGATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 7191
QY 961 AGGCAAAAAAAGCTTGTCTAGCCCAAAACAGATGCAAGGATGATGTTTAAAAATGG 1020
DB A-----AAGAAAGATTTGAGGACCAAAATCAGATGCA-----CGGTGAATGTTCAAGAT 7251
QY 1021 GAAAAAGATCTTAAATTTATCTGCAATTTGACATTTTATGACAAAAAACAAGTTTAAAGC 1080
DB GAAATTAAGATCTCAATCTCTATCTGCTGTTTATTTG--AAAAACAGATCTGAGC 7302
QY 1081 AGTTACAGATTTTATATGTTTATTTATTTGCGATATGATGATGATGATGATGATGATG 1140
DB AACTGCTCATGTTTATCAATATGTTGATGCGGATATGATGTTGTTGGAATCTGCCAC 7361
QY 1141 CTACAGTTGATGATTTGAAAAATATATCTCTGCTCAAAAGAGCATGATGATTTAAAC 1200
DB CAACTGTGATGATTTGAAAAATATATACCAAGTCAAAAAACGCTGCTGATGATGATG 7421

QY 1201 CAGGATTTACAGATCTCTGGCAGGTTATGCTGCTGATTAATATACAGACTTTCAGCAGC 1260
DB CAGAAATTTACAGGCTTATGCGACGCAAGTGTGATGATTAATATACCACTTGTATGAGG 7481
QY 1261 TAGTTGGTTGACTTATGATATCATTTGATTAATTTGATCTATCTGCTGATTTTAAATTT 1320
DB TGGTAAAGACTGATGTTGAAATATGATTAATCTGCTGATTTGATGCTGATTAACAGATAT 7541
QY 1321 TATTTAAAGCAGTAAAGTTGATTTGTTGAGAGGGAATTAAGT----- 1365
DB TGTGTAAGACTATTTGCTGCTGTTTATTAATAAGAGGGAATGATGATGAGCCCTCTTCA 7601
QY 1366 -----AAGATATAGAAAGTTTGTGCTGCTTCTTCAAGGAGGACATTTG 1412
DB TTCTGTATTTGAGAGAAAAATGAAAAATTTGTTTATGATGATGATGATGAGGACATTTA 7661
QY 1413 ACTGACTGTTATTTGTTTAAACCGTTTGGAGAGAGAACGTTTGTGGTAACTTT 1472
DB GCACATTTAAATATGTTTAAACCATTTCTGAGAGAGTAAAGAACGTTTGTGGTAACTTT 7721
QY 1473 GATTAAGAGATGCAAGATCTTTTGAAGATTAAGAAAAATGATCAATGTTACTTTCA 1532
DB GATTAAGAGATGCAAGATCTTTTAAAGATTAAGAAAAATGATCAATGTTACTTTCT 7781
QY 1533 ACAATTCGCAATCTCATTAATTTAGTAAATAATCTTCTTACCTTTCAAAATTTTACGT 1592
DB ACATTCGCAATCTTAATTAATTTATTAATTAATTTATTTCTTCTTCTTCTTCTTCTTCT 7841
QY 1593 GATGAGAACCAAGATTTATTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1652
DB AAAGAAAAACCTGATGCTATTTATTTATCTGAGAGCTGCTGAGCTGCTGCTGCTGCTTAC 7901
QY 1653 ATGGAATCTATTTGAGCAAGAGATTTATTAAGATTTGATGATGATTTGATGATTTAA 1712
DB ATTTGATTAATCTATTTGAGCTAAAGAGCTTTATTAATGATCTTTGATTAATGATTTAA 7961
QY 1713 TCTACATTAATCTGAAAAATAGTTTATCCGTAACAGATTTTATTTATTTGTTCAAGTGA 1772
DB CCAACAGTTTACAGAAAAATTTATGTTATCTGATCCCTGATCGATTAAGTTCTTCAATGGA 8021
QY 1773 GAAATGAAGAGTATATCTTAATCTATTTATTTACTTGGGAGTATTTTATTAAGATTTT 1832
DB GAGATGAAAAATGTTATCTCAAAAGCTATTTATTTGAGGAGTATCTTTTATGATTTT 8081
QY 1833 TAACTAGTAACTCATGAACAAAGTTTATGATTTGATTAAGATGATTTATTTGA 1892
DB TAACTAGTAACTCATGAACAAAGTTTATGATTTGATTTGATTTGATTTGATTTGATTTG 8141
QY 1893 AAAAAATGAGATTAACCGACGAATATTTATTTCAACAGATATTTCTGATATTT 1952
DB TTGAAACAGAGATTAAGAGAGAGTCTTTATTTCAACAGATATTTCTATTTAGAAC 8201
QY 1953 CAGATATTTGCAAGTATTAATAATTTCTCAGTTACAAAGATGAGAACATATTTAAACA 2012
DB CTAAATTTTGTGATGCTGCTGCTGTTGATTAAGTTTATGATTAAGATTAATTTATGAAA 8261
QY 2013 AATCAGATGATTTATTTGCAAGAGGCCCCCTTCTTATTTATGATTTCAATTTCAAG 2072
DB AAGCTATATTTATTTATCTACGACGAGGCTCCAGCTTTTATGAGGCTCATTAACAT 8321
QY 2073 GAAAAAACAATTAATTTTCTTCAAGAAAAAAGTATGTTGAACATGTAATGATCATC 2132
DB GCAAAAGCCAAATTTGTTGCTTCAAGAAAAAGTATGTTGAACATGTAATGATCATC 8381
QY 2133 AAGTATGTTTGTAAAGAAATTTTCAAGATTAATTAATTTTATTTATGAAAAATATAG 2192
DB AAGTATGTTTGTAAAGAAATTTGCAAGAAATTTGCAAGAAATTTGCAAGTATTTCTGTTGAAGAT 8441
QY 2193 ATGATTTGTTGAAAAATTTATTTGATTTCTTA 2226
DB TAGATTAATTTGGGAGACTATTTGAAAAATTTACAA 8501

RESULT 5
STH289861 21562 bp DNA linear BCT 15-APR-2005
LOCUS Streptococcus thermophilus eps locus.
DEFINITION
ACCESSION AJ289861
VERSION GI:18076389
KEYWORDS deod gene; Deod protein; Eps 2; Eps 3; Eps 4; eps10 gene; Eps10
protein; eps11 gene; Eps11 protein; eps12 gene; Eps12 protein;
eps14 gene; Eps14 protein; eps15 gene; Eps15 protein; eps16 gene;
Eps16 protein; eps2 gene; eps3 gene; eps4 gene; eps5 gene; Eps5
protein; eps9 gene; Eps9 protein; epsa gene; Epsa protein; epsB
gene; EpsB protein; epsC gene; EpsC protein; epsd gene; Epsd
protein; epsB gene; EpsE protein; ORF1; ORF6; pseudogene.
SOURCE Streptococcus thermophilus
ORGANISM Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Pluvinet, A., Charron-Bourgoin, F., Morel, C. and Decaris, B.
TITLE Implication of horizontal transfers in the chimERIC structure of
the eps locus of Streptococcus thermophilus IP6757
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 21562)
AUTHORS Pluvinet, A.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-2000) Pluvinet A., Laboratoire de Genetique et
Microbiologie, Université Henri Poincaré - Faculté des Sciences,
B.P. 239, Vandœuvre-lès-Nancy 54506, FRANCE
FEATURES
source
1. 21562
/organism="Streptococcus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:1308"
<1..465
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC82000.1"
/db_xref="GI:18076390"
/db_xref="UniProt/TrEMBL:Q8VM12"
/translation="DIYIVGCGAQLLSVAYGCKEHSPEIREIGVPEPLTLTDEGLAD
EHTKDFGSLTNGMHGDMGSLSDKAVLATSKGYPHQIICFSPKHYAFQAHLEPDL
AIDLLTADESEHLVQONKQLDFVQPEQAHNDYSQMKKLFALDSLTLQI"
800..1809
/gene="deod"
/gene="deod"
800..1510
/gene="deod"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Deod protein"
/protein_id="CAC82001.1"
/db_xref="GI:18076391"
/db_xref="GOA:Q8VM11"
/db_xref="InterPro:IPR000845"
/db_xref="InterPro:IPR004402"
/db_xref="UniProt/TrEMBL:Q8VM11"
/translation="MSIHISAKQGIADKILLPDPPIRAKFIATNLEDAVCNEVEN
MFGVGTGKSRISVMGTGMPGISIYARELLVDYGVKKLIRVGTAGSLNENVARE
LVLAQAATNSNIIRNDMPQYDFQIANFNLLDKAYHIAKNGFTTHAGVLSDDVY
SNYERKNIELGKGVKAVEMEAALYYLAQHQVDALAMTISLSLVNPDDEDTJAER
QNTFTDMKKGLETLIAD"
1744..1764
/gene="deod"
/gene="deod"
1781..1786
/gene="deod"
1804..1809
/gene="deod"
1853..3307
/gene="epsa"
1853..3307
/gene="epsa"
CDS

/function="putative exopolysaccharide regulation"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Epsa protein"
/protein_id="CAC82002.1"
/db_xref="GI:18076392"
/db_xref="GOA:Q56038"
/db_xref="InterPro:IPR004190"
/db_xref="InterPro:IPR004474"
/db_xref="UniProt/TrEMBL:Q56038"
/translation="MSSRTNRKQKHTNSGCMGVNNGVLTIIYAILALVLLFTPMYNF
LSRFLNIIITIGLVLAIALSIFLOKTKKPLVTVVAVIVISLVISLVGFGKRMIDI
TRNMQGTAAFSSEVMSIVPKESDIDKVSQLTYSVAPTKVDKNIEILMSALKKKKV
DKVDVDAVYQEAVDNLKSGSKAMVLSGSVSLSEVDSNYSNLKTIYTKIKKN
NSANQVDSRVFNIIYISGIDTYGPISYRSQDVNIIMVNMNTHRLITLTPRDAVYK
IPGGADQYDKLTTHAGIYGVETSEQTEBLVGIKLDYARINFTSFLKIDLGCVTV
HNDQAFPOEKPEPVPYDIOVNSQEAALGFRERYNIDGGNDGRKQOEKVISAILKKA
SUKSVNFTSIYNNIQDSYQTMSLNTINALANTQLEBSKRTYVSQAVTGTGTQGL
ISYAMPNSSLVYMKLNDSSVESASQAIKKLMEEK"
3308..4039
/gene="epsB"
/gene="epsB"
3308..4039
/gene="epsB"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="EpsB protein"
/protein_id="CAC82003.1"
/db_xref="GI:18076393"
/db_xref="GOA:Q56039"
/db_xref="InterPro:IPR004013"
/db_xref="UniProt/TrEMBL:Q56039"
/translation="MIDVASHIVPVDVDEPTELESIDLIGESYAQGVKRIYSTSHR
KCMFERPDKIRPANKFKVKAABALYPLDITYYGELYYTSDIVEKLEKNLIPRMNT
QNALIFRSARKTWEKHSGLSNVLRAAGVPIYAHIERVYALENMDRVEIIMGCYT
QVNSHVLPKLFQGDQKVRKGRVRFLEKNIVHVASDMNLGRRPPMKDAYEIVK
KNYGSKRANLFIENPKTILENOYL"
4040..4740
/gene="epsC"
/gene="epsC"
4048..4740
/gene="epsC"
/function="chain length determinant and exporter"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="EpsC protein"
/protein_id="CAC82004.1"
/db_xref="GI:18076394"
/db_xref="GOA:Q8VM10"
/db_xref="InterPro:IPR003856"
/db_xref="InterPro:IPR005701"
/db_xref="InterPro:IPR005829"
/db_xref="UniProt/TrEMBL:Q8VM10"
/translation="MNQDNTKSDSIDVLLHLKLTWKLLILFTAYPAVPSFLGYF
FIQPLTYSTRIRIVVQAQDNKLSQDQAGTYLANDYKEIITSNDVASEVLDKEL
NLSBALSKMVSVINIPTDRLTISIVNATGQDAQTLANKVREAVASKIKKIVKEDV
TTLSEKKLPESPSPNIKLVLLGAVLGFLAVGVLVREILLDRVRREDVDALGM
TLGIYPTDKIT"
4750..5490
/gene="epsD"
/gene="epsD"
4750..5490
/gene="epsD"
/function="chain length determinant and exporter"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Epsd protein"
/protein_id="CAC82005.1"
/db_xref="GI:18076395"
/db_xref="GOA:Q8VM09"
/db_xref="InterPro:IPR005702"
/db_xref="UniProt/TrEMBL:Q8VM09"

Db	8348	AGCACTTAAATTTTAAAGAAAAAGAAAGCCTGACGTTTATTATTCATCAGAGCAGCGGT	8407
Qy	1634	TGCTGTCCCTCTTCTTTACATCGGAAACTATTTGGAGCAAGAAGATTTATTTGAAGT	1693
Db	8408	AGCAGTTCATTTTTTTTACTAGTTAACTATTTGGAGCGAAAACGGTTTATATAGAAGT	8467
Qy	1694	ATTGGATCGAGTTATTAATATCTACCTTTACTGGAAGAACTAGTTTATCCCGTAAACAGTAT	1753
Db	8468	ATTGATATGAATAGATTAAACGACACGTAACTCGGAAATGGTTTATTCAGTGAACAGTAA	8527
Qy	1754	TTTTATTTCTGACGGGGAAGAAATGAAGAAGTATATCTTAATCTTAACTCTGGGGAG	1813
Db	8528	ATTATATGTTCAAGGGAGAGAGATGAATAACTGTATATCCAAAGCTATTAATCTGGGGAG	8587
Qy	1814	TATTTTAAATGATTTTGTGTAACAGTAGAAGCTCATGAACAACAGTTTAATCGATTGATA	1873
Db	8588	TATTTTAAATGATTTTGTGTTACAGTGGGAACCCAGAACGCCCTTAAATAGGCTTATT	8647
Qy	1874	AAAGAGATTGATTATTTGAAAAAAATGAAAGTATACCGACGAATATTTTATTCAAACA	1933
Db	8648	AAGGAAGTTGATCGTTTAAAAAAGAAAGGTATTTATTAACAGATGAGGTTTTTATTCAGACA	8707
Qy	1934	GGATATTTCTGACTATATTCGAGAAATTTGCAAGTTATTAATAATTTCTCAGTTACAAAGAA	1993
Db	8708	GGTTTTTCAACTTAAGAACTTCMAATACGTGACGTGAAAAACATTATTTCTTATCTGAA	8767
Qy	1994	ATGGAACATATTTATTAACAAATCAGAAAGTAGTTATTTGCCACGGAGCCCCGCTACTTTT	2053
Db	8768	ATGGAAAGTTATATGATATCGTGCAGATATTTATTTACAGCATGTGTGTCTCAGCCACATTC	8827
Qy	2054	ATGAATTCATTATCCAAAGAAAAAAACAAATTAATGTTTCTGTAGCAAAAAAAGTATGCT	2113
Db	8828	ATGGAGCAATGCTTAAAGAAAAAAACCGATGTTGTTCCAAAGCAGAAAAAGTTTGA	8887
Qy	2114	GAACATGTAATGATCATCAGTAGAGTTTGTAAAGAAATTTTCA--AGATTAATAT	2170
Db	8888	GAGCATGTAAATGATCACCAATTAATTTTGCTATCAAGTTAAAGATAGGTACGATAT	8947
Qy	2171	ATTTTATTTATGAAGAAATATAGATGATTTGTTGAAAAAATTATGAAGTTCTTAAGCA	2230
Db	8948	ATTGAAGTTGTGGAAGATATCAAACTTTACACAATTTTGAACAACAGATTTTAAGTATA	9007
Qy	2231	ACTACTTTTACA---TCAATATATATTTTTTTTGTGAAAGTTTAAACAATAGTTGAA	2287
Db	9008	TCAGAGATGACAACTTCATATATATAGAATTTATTAAGAACAACTTAGCAAGAAGTTG-A	9066
Qy	2288	AAATTTAATGAGATCAAGAAAATGAATATATAAAAAGATGCATATTTGATATGCTTAA	2347
Db	9067	AAATATTAATGGGTAAAGTAAACTACTTTTAAATATATATATTAAGAAGATGATGTATAGA	9126
Qy	2348	TCATATATTTTCTCAGATTTTACTGGAAGGGATACAGATA	2388
Db	9127	ATTATCATCAAGTGTATTTCTATGAAAGTTTAAAGTATATA	9167

RESULT 6
CP000023_09/c
WPCOMMENT
Sequence split into 18 fragments LOCUS CP000023 Accession CP000023

Fragment Name	Begin	End	LOCUS	CP000023	Accession	CP000023
CP000023_00	1	110000				
CP000023_01	100001	210000				
CP000023_02	200001	310000				
CP000023_03	300001	410000				
CP000023_04	400001	510000				
CP000023_05	500001	610000				
CP000023_06	600001	710000				
CP000023_07	700001	810000				
CP000023_08	800001	910000				
CP000023_09	900001	1010000				
CP000023_10	1000001	1110000				
CP000023_11	1100001	1210000				

CP000023_12	1200001	1310000
CP000023_13	1300001	1410000
CP000023_14	1400001	1510000
CP000023_15	1500001	1610000
CP000023_16	1600001	1710000
CP000023_17	1700001	1796846
Continuation (10 of 18) of CP000023 From base 900001 (CP000023 Streptococcus thermophilu		
Query Match	12.8%;	Score 892.6; DB 1;
Best Local Similarity	66.9%;	Pred. No. 1.2e-81;
Matches 1393; Conservative	0;	Mismatches 669; Indels 19; Gaps 8;

QY	317	ATACCTATTAAAGGTAAATTAAGAAGTTTAAAGATAGCTTTCATTTTGCACAACTCTA	376
Db	77023	ATACTTTATCAACTGCACGTGGCGTATGATACAAATATCTATTCCTTATGACCTGCTGAACA	76964
QY	377	TCAAAAAGAACGATTTCTAATTTCACACGGCTGACGATGGGAAAAATATGCAATTTTAT	436
Db	76963	ACGAGAAATATACATTTGTGATTACACCACTGAGAGATTGGCATTAATGGAGGGGCTCTT	76904
QY	437	TGAATCATATPAAACAAATTCAAAAAATCTTGTTCATCTGGTATGTTTAAGTACGAAT	496
Db	76903	TGAGTCAGATCAACTTCTCTATCAAAATATTTGGACGATATAGTATGATCTGGAGATGTGA	76844
QY	497	AGATPAAATTAATTTATTCATTACCGCTATATTCTGTGGAAAGACATATGAGTTTTC	556
Db	76843	AGTGGCGTTTCCAGAGGAGTCCCAATPAAATTCCTTTT--GATGATGCGATTGACTTTGC	76787
QY	557	AACAAGGAATGAGTGCACACAGCTTTTAAATCTAACCAAGTGAATTTTAAACGTAAA	616
Db	76786	GATCTACGAAGTTGTGACCATGTGTTTTCAACTTACCGAGTGAACATTAAGATCTCAA	76727
QY	617	GCAATTCGTTTCAGATTTTGAATTTGATTGATGATGTAAGCGTTGATTTATTTCAAT	676
Db	76726	ACATCTTGTTCCGATTTTGAATGATCGGATGATGATGATGATGATGATTAATTAATTAAT	76667
QY	677	CGGTTTATCTGCGTTGAAAAACAAAAATCCAAACGCTGAGGTGACATAGCAATTTGTAC	736
Db	76666	TGATTTTCAAGGCTTTTAAAAATATAAAAAATCAACAAAGTTGAGACATATGATGTGAC	76607
QY	737	TTTTTCCAAATTTTATATAGCCTATGCTATCATGATGAAACGACTTTTGATATACT	796
Db	76606	TTTTTATCTCCATTTACTACAAATATAGCCATATCTTTCAAAAGCGATGTGGATATCTT	76547
QY	797	CGAGCGGATCGGGTTAATTAATTTGTGATATGTTTCAATTTTGTATGTTTCCAAATT	856
Db	76546	TGGGCGTGTGATGTGCTTCTAATTTGTGGTCTGTGCGGAAATGTCTTATGCTTCATCAT	76487
QY	857	TCTATAGATATGTGTG-AACGGCTATTTTTGCTCAGAAAGAGGTGGACGAATGACGCA	915
Db	76486	TCCGAAAGATGAGGAACCTGTCTAATTTTCTGTAGAAACGAGTAGGAAAAACGGTCTGA	76427
QY	916	TATTTACATTTACAGATTTTCAGATCGATGTATGTTATGCTGTGAGACGCAAAAAAGCT	975
Db	76426	TCTTTAATTTTATTAATTCGTTCTATATGATATATATGTCAGAAAAACGAAAAAAGAAC	76367
QY	976	TGCTCAGCCAAAACAGATGCAAGGATGGGTATGTTTAAATGGAAAAACGATCTTAG	1035
Db	76366	TTATGGCTCAAAAACCAATGCA-----GGGTGTATGTTTAAATGATTAATGACCAACG	76312
QY	1036	AATTAATTCGAATTTGACATTTTCATAGCCAAAAACAGTTTATAGACGATTTACCAAGTTT	1095
Db	76311	GATTACACCAATTTGTCAGTTTATCCG-AAAAACAAGTCTAGATGATTTCCACAGTTT	76253
QY	1096	ATAATGTTTAAATTTGGCGATATGATGTATGTTGTGTACAGTCCACCTACAGTTGATGAAT	1155
Db	76252	ATAATGTATCTATGTTGGGATATATGATGCTTGTATGATCTGTCCACCGACAGTGAATGAT	76199
QY	1156	TTGAAAAATATATCTCTGTCAAAAAGACGATTTGATGATTTTAAACCAAGGATTTACAGTTC	1215
Db	76192	TTGAAAAGTATACACCAAGTCAAAAACGTGGCTCAGCTTCAAGCCAGAAATTAACAGTTC	76133
QY	1216	TCGCGACGTTAGTGTGTGATATATTCACAGACTTTCACACGACGTATGTTGGTGGACT	1275


```

Db      76132 TATGACAGTAAGTGTCTAGCAATATTAACAATTGAGAGTGTAAAGCTCATG 76073
Qy      1276 TAGATACATTTGATTAATGACATCTGTCAGATATTAATAATTTATTAAGCACTGA 1335
Db      76072 TTGAATATATCGAATACGTGCGATTTGTCGATATCAAGATTTGTGAAAGCTATCG 76013
Qy      1336 AAGTGTATTTGTCAGAGAGGAAGTAAGTAAA--AGTATATGAAGTTGTTGTCGG 1393
Db      76012 TCCGGGTTTTCAAAAAGAGGAGATAGTAGATTAATATATGAAAGTTGTTAGTAGG 75953
Qy      1394 TTCTTCAGGGGACATTTGACTCACTTGATTTGTTAAACCGTTTGAAGAGAGAGA 1453
Db      75952 TTCTTCGATGGGACATTTGGACACTTGAAATATCTTAAACCTTTTGAAGTGAACGA 75893
Qy      1454 ACGTTTGGGTACATTTTGATTAAGAGATGCAAGAGCTTTTGAAGATGAAAAAT 1513
Db      75892 CCGCTTTGGGTACCTTTGATTAAGAGATGCAAGAGATTTTAAAGATGACAGTT 75833
Qy      1514 GTATCCATGTTACTTCACAACAAATCGCAATCTCAATTAATTTAGTGAATACTTCTT 1573
Db      75832 TTATCCGTCTATTTCTCTACTACAGAAATTTTAAATTTAGTAAAGATCTTTT 75773
Qy      1574 AGCTTCAAAATTTTACGTGTAGAGAAACAGATGTTATTTATTCATCTGTCGCGCT 1633
Db      75772 AGCACTAAATTTTAAAGAAAGAAAGCCGACGTTATTTATTCATCAGAGACGCGT 75713
Qy      1634 TGTCTGCCCTCTCTTTTACATCGGAAAATTTTGAAGCAAGACATTTATATGAAGT 1693
Db      75712 AGCAGTCCATTTTAACTTACCTAGGTAAATTTTGAAGCAAGAAACGTTTATATAGAAT 75653
Qy      1694 ATTGATCGAGTTAATTAATCTACATTAACCTGAAAACCTAGTTATCCCTGAACAGATAT 1753
Db      75652 ATTGATGAATGATGATTAACCGACATGACGGAATGGTTATCTCAGTGAACAGATTA 75593
Qy      1754 TTTTATGTTCACTGGGAGAAATGAAGAGTATCTTAAATCTATTAACCTGGGAG 1813
Db      75592 ATTATATGTTCACTGGGAGAGATGAAGAACTGTCTATCCAAAGCTATTAATCTGGGAG 75533
Qy      1814 TATTTTAAATGATTTTGTGTAACAGTAGAACTCATGAACAACGTTTAAATCGATTGATA 1873
Db      75532 TATTTTAAATGATTTTGTGTAACAGTAGAACTCATGAACAACGTTTAAATAGGCTTAT 75473
Qy      1874 AAAGAGTTGATTTATGAAAAAAATGGAAGTTAACCGAAGAAATTTTATTAACA 1933
Db      75472 AAGGAGTTGATCGTTTAAAAAAAGAGATTAATTAACAAGTGGTTTATTAACAACA 75413
Qy      1934 GATATATCTGACTATATTTCCAGAAATTTGCAAGTATTAATAAATTTCTCAGTTACAAGAA 1993
Db      75412 GATTTTCAACTTATGAACTCAATCTGATGCTGAGAAACATTAATTTCTATCTGAA 75353
Qy      1994 ATGGAACATATATTAACAATCGAAGTATTTTGAACGAGAGCCCGCTACTTTT 2053
Db      75352 ATGGAACATATATGATCGAGATCGAGATTAATTAATCAAGATGTTGTCAGACATTC 75293
Qy      2054 ATGAATTCATTTCCAAAGAGAAAAAACAATTAATTTTCTTGAACAAAAAGTATGCT 2113
Db      75292 ATGGAACATATGCTAAAGAGAAAAACCAATGTTGTTCCAAACAGGAAAAAGTTTGA 75233
Qy      2114 GAACATGTAATGATCATCAAGTATGTTGTAAGAGAAATTTTACA--AGATTAATAT 2170
Db      75232 GAGATATTAATGATCAACCAATTTGATTTGCTATCAAGTTAAAGTAGTACATAT 75173
Qy      2171 ATTTTATTAATGAAAAATATAGATTTGTTGAAAAAAATTAATGAAGTTCTAAGCAA 2230
Db      75172 ATGGAAGTTGTCAGAGATATCAAAACCTTACAACATTTTGAACAAGATTAATAGATA 75113
Qy      2231 ACTTACTTACA--TCAATTAATATTTTGTGAAAGATTAACAACATATGTTGA 2287
Db      75112 TCGAGAGTACACTTAAATTAATTAAGAAATTTATATACAACTTACAGAAAGTTG-A 75054
Qy      2288 AAATTAATGAGATCAAGAAATGATTAATTAATAAAGATGATTTGATTAATGCTTA 2347

```

```

Db      75053 AAATTAATGAGGTAAAGTAAACTACTTTTAAATATATTAAGAAATGATTTGATA 74994
Qy      2348 TCATTAATTTTCTCAGATTTTACGAGAGAGATACAGATA 2388
Db      74993 AATTACATCAAGTGTATTTCTATGAAGTTTAAATATATA 74953

RESULT 7
CO963297
LOCUS      CO963297 16032 bp DNA linear PAT 06-DEC-2004
DEFINITION Sequence 1 from Patent WO2004085607.
ACCESSION CO963297
VERSION    CO963297.1 GI:56403543
KEYWORDS
SOURCE
ORGANISM  Streptococcus thermophilus
            Streptococcus thermophilus
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.

REFERENCE
1 Horvath, P., Manoury, E., Huppert, S. and Fremaux, C.
  Texturizing lactic bacteria
  Patent: WO 2004085607-A 1 07-OCT-2004;
  Rhodia Chimie (FR)

FEATURES
    source
        1..16032
        /organism="Streptococcus thermophilus"
        /mol_type="unassigned DNA"
        /db_xref="taxon:1308"

ORIGIN
Query Match 12.7%; Score 889; DB 6; Length 16032;
Best Local Similarity 6.8%; Pred. No. 6,7e-81;
Matches 1344; Conservative 0; Mismatches 655; Indels 12; Gaps 5;

Qy      317 ATACCTAATTTACGTAATTAATTAAGCAATTAAAGATAGCTTTCTATTTTCAACATCTA 376
Db      5592 ATACTTATATCACTGCACTGTGCGTAGATCAAAATATCTATTCCTTATGACTGCTGACA 5651
Qy      377 TCAAAAAAGACATTTCTAATTAACAACGCTGAACATGAGAAATATGCAAGTTTAT 436
Db      5652 ACAGAAAGATACATTTGATGATTAACAACACTGAGAGATTCGATTAATGAGAGGCTCTT 5711
Qy      437 TGAATCACTAATCAAAATTCAAAAAATCTGTTGCACTGGTAGTTTATGATCAGAAAT 496
Db      5712 TGAGTCAGATCAACTTCTATCAAAATTTTGGCAGATATGATGCTATCGAATGCTGA 5771
Qy      497 AGATTAATTAATTTATTCATTAACGCTCTAATTAATCTGTGGAAGAAAGCTATAGATTTC 556
Db      5772 AGTGGCGTTTCCAGAGGAGTCCCAATTAATTCCTTT--GATGATGCAATGACTTTCG 5828
Qy      557 AACAGGGAAGTGTGACCACTCTTTATTAATCTACCAAGTGAAGTTTATAGACGTAA 616
Db      5829 GACTCATGAAGTGTGACCATGTGTTATCAACTTACCAAGGAAACATTAAGCATTCGAA 5888
Qy      617 GCAATTCGTTTCAGATTTTGAAGTTGATGATTAATGTAAGGCTGATTAATTCAT 676
Db      5889 ACATCTGTTTCGATTTTGAAGTCAATGATTAATGATGATGATTAATTAATTAAT 5948
Qy      677 CGGTTTACTGCGTTGAAAAAACAATACTCAACTGCTAGTGAACCATAGATTGTAAC 736
Db      5949 TGATTTCAAGGCTTTAAAAAATTAATAAATCAAAAGTGGAGACATATGATTCGTGAC 6008
Qy      737 TTTTTCACAATTTTATTAAGCTTATCATATCATATGAGAAACGACTTTTGATATACT 796
Db      6009 TTTTAACTCCAAATTTACTACAAACATGACATATCTTTCTAAACGCAATGTTGATATCTT 6068
Qy      797 CGAGCGGTAGTGGGTTAATTTTGTGATAGTTTCTATTTGTTAGTTCCAAATAT 856
Db      6069 TGGGGCTGTATTTGCTTCTAATTTGTGTCTGTGCGGATTTGTCTTACTCTATCAT 6128
Qy      857 TCGTAGAGATGTGG-ACCGGCTAATTTTGTGCAAGACAGTTGACAGAAATGACGCA 915
Db      6129 TCGGAAAGATGAGGAACCTGATTTTGTTCAGAAACAGATGAGGAAAAACGGTCTGA 6188

```

QY 916 TATTTCATTTCTACAGTTTCGATCGATGTATGTGATGCTGAGAGCCGCAAAAAAGACT 975
 |||||
 Db 6189 TCTTTAAATTTTATTAATTCGTTCTATGTATATAGATGAGAAAAACGGAAAAAGAAC 6248
 QY 976 TCGTCAGCCAAAACGATGCAAGGGGTATGTTTAAATGGAAAAACGATCTTAG 1035
 |||||
 Db 6249 TATAGCTCAAAAACGATGCA-----GGGTGTATGTTTAAATGATATATACCAAG 6303
 QY 1036 AATTACTCCAAATGGACATTTCTATGCGAAAAACAAGTTTGAAGAGTTTACGAGTTT 1095
 |||||
 Db 6304 GATTACACCAATTTGTCAGTTTATCCG-AAAAACAAGTCTGATGTAATGCGACAGTTT 6362
 QY 1096 ATAAATGTTTAAATTTGCGATATGAGTCTAGTGTATACAGTCCACCTACAGTTGATGAT 1155
 |||||
 Db 6363 ATATATGATCTAGTGTGATGATATAGCTGTAGTACTGCTCCACGACAGTGTAGTAAT 6422
 QY 1156 TTGAAAAATATATCTCTGTCAAAAGAGACGATTTAGTTTAAACCAAGGATTTACAGTTC 1215
 |||||
 Db 6423 TTGAAAAATATATCAACCAAGTCAAAAAACGACCTTAATTTTAAACCGGGATTTACCGGCT 6482
 QY 1216 TCTGCGAGGTTAGTGGTCGTAGTAATATCAAGACTTCGACGAGTATGTTGGTGGACT 1275
 |||||
 Db 6483 TATGCGAAGTTATGTCGCGGTAGCAATATACGACCTTTGACGAAGTATGTTAACTGATG 6542
 QY 1276 TACGATACATTTGATTAATTTGACTATCTGTCAGATATTTAAATTTTAAAGACAGTGA 1335
 |||||
 Db 6543 TTGAATATACATATGATATCTGTCATTTGTCGATATCAAGATATTTGTTGAAGCTATCT 6602
 QY 1336 AAGTTGATTTGTTGAGAGAGGAAATGATGTAAGTAA-AGTATATGAAAGTTTGTTCGCG 1393
 |||||
 Db 6603 TCGTAGATTTTAAAAAAAGAGGAAATGATGATATATGATGAAAGTTTGTTCGAGG 6662
 QY 1394 TTCTTCAGGGGACATTTGACTGATTTTGTATTAACCGTTTGGAAAGAGAGA 1453
 |||||
 Db 6663 TTCTTCGTCGCGCATTTGGACACTGATATGCTTAAACCTTTTGGAGTGAACGGA 6722
 QY 1454 AGCTTTTGGGTAAATTTGATTAAGAGATGCAAGAGCTTTTGAAGATGAAAAAT 1513
 |||||
 Db 6723 CCGCTTTTGGGTATCTTTGATTAAGAGATGCAAGAGATTTTAAAGATGACGAGT 6782
 QY 1514 GTATTCATGTTTACTCTTCAACAAATGCAATCTCATTAAATTTAGTGAATAATCTTTCTT 1573
 |||||
 Db 6783 TTATCGTGTATTTCTTCTACTAAACAGAAATTTTAAATTTAGTAAAGAAATCTTTT 6842
 QY 1574 AGCTTTCAAAATTTTACGTATGAGAAACCAAGATTTATTTCACTGTGCGCGCT 1633
 |||||
 Db 6843 AGCACTTAAATTTTAAAGAAAAAGAAAGACCTGACGTTATTTTATCATCGAGAGCGGT 6902
 QY 1634 TGCTGTCCCTCTCTTTTACATCGGAAACCTATTTGGAGCAAAAGACGATTTATTTGAAGT 1693
 |||||
 Db 6903 AGCAGTTCGCTTTTATTTATCTGTAAACCTTTGGAGCAAAACGTTTATATNAGAGT 6962
 QY 1694 ATTTGATCGATTAATTAATCTACATTTACTGAAAACTAGTTTATCCCGTACAGATAT 1753
 |||||
 Db 6963 ATTTGATGAAATGATTAACCGACAGTAACCTGGAATTTGGTTTATCCAGTGAACAGATAA 7022
 QY 1754 TTTTATTTGTCAGTGGAGAAATGAAAGAGTATATCTTAAATCTTAACTTTGGGAG 1813
 |||||
 Db 7023 ATTTATTTGTCAGTGGAGGAGATGAAACGTCTATCTCCAAACCTATTTATCTGGGAG 7082
 QY 1814 TATTTTAAATGATTTTGTGTAACAGTAGAACTCATGAAACAAGTTTAAATCATGATGATA 1873
 |||||
 Db 7083 TATTTTAAATGATTTTGTGTAACAGTGGAAACCAAGAACGCTTTTAAATAGCTTATTT 7142
 QY 1874 AAAGAGATGATTTTATGAAAAAAATGGAAGTATTAACCGAGGAATTTTATTTCAACA 1933
 |||||
 Db 7143 AAAGAGATGATTTTATGAAAAAAAGAGTATTTTATTAACAGTGAAGTTTATTTATCAACA 7202
 QY 1934 GGATATTTCTAGCTATATTTCCGAATTTTGCAGATATTAATAAATTTTTCAGTTACAAAGAA 1993
 |||||
 Db 7203 GGTTTTCAACTATGAAACCTCAATACTGTACTGAAAAAAACTTATCTTATATTCGAG 7262

QY 1994 ATGGAACAATATATTTAACAATCAGAGTATTTTGCACGAGAGCCCGCTACTTT 2053
 |||||
 Db 7263 ATGGAACAATATATTTAACAATCAGAGTATTTTGCACGAGAGCCCGCTACTTT 7322
 QY 2054 ATGGAATTCATTTATCCAAAGAAAAAAACATATTTGTTTCTTACGACAAAAAAGTATGCT 2113
 |||||
 Db 7323 ATGGAAGCATTTGTCAAAGGCAAAAAACCAATTTGTTTCTTCAAGCAAGAAAAATATGGA 7382
 QY 2114 GAACATGTAATATGATCATCACTAGTATGTTTGAAGAAAGATTTTCAAGATTAATATTT 2173
 |||||
 Db 7383 GAACATGTAATATGATCATCACTAGTATGTTTGAAGAAAGATTTTCAAGATTAATATTT 2233
 QY 2174 TTAATTTATGAAATATATGATATTTGTTTGAAGAAATATATGAGTTTCTAAGCAAACT 2293
 |||||
 Db 7443 ATTTATTTGTTGTAAGATATTTACTCTGCTGTAATTTTATTAAGAAAGTATGATCTTT 7502
 QY 2234 AACTTTACATCAAAATATATTTTGTGTAAGATTTTAAACAAATAGTTGAAAAATTT 2293
 |||||
 Db 7503 ATTTGCAAAAGCAGATGATATTTTACCTTAAAGATGACAGTTTATTTTAAATTTCAAA 7562
 QY 2294 AATGAGATCAAGAAAAATGATATATTAATAA 2324
 Db 7563 GAGATGTTGATGAACCTGTAGAGAGAGA 7593
 RESULT 8
 CR931662
 LOCUS
 DEFINITION
 Streptococcus pneumoniae strain 34359 (serotype 14).
 CR931662
 VERSION
 CR931662.1 GI:68642995
 KEYWORDS
 Streptococcus pneumoniae
 Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus
 ORGANISM
 1 (bases 1 to 19918)
 Bentley,S.D., Aarensen,D., Mavroidi,A., Saunders,D.,
 Rabinowitch,E., Collins,M., Donaghy,K., Harris,D., Kalofic,M.S.,
 Murphy,L., Quail,M.A., Samuel,G., Skovsted,I.C., Barrell,B.G.,
 Reeves,P., Parkhill,J. and Spratt,B.G.
 Genetic analysis of the capsular biosynthetic locus from all 90
 serotypes of Streptococcus pneumoniae
 Unpublished
 2 (bases 1 to 19918)
 Bentley,S.D.
 Direct Submission
 Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
 Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, E-mail: sds@sanger.ac.uk
 NOTE: This sequence was generated from a PCR product representing
 the region from dexB to a1a and is not necessarily responsible for
 the expressed capsule serotype. For a detailed description of how
 CDS products were predicted see the associated publication.
 CDS products were predicted see the associated publication.
 FEATURES
 source
 1.19918
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /strain="34359"
 /db_xref="taxon:1313"
 <1.134
 /gene="dexB"
 /locus_tag="SPC14_0001"
 <1.134
 /gene="dexB"
 /locus_tag="SPC14_0001"
 /codon_start=3
 /transl_table=11
 /protein_id="CAI33314.1"
 /db_xref="GI:68642996"
 /translation="SNBBDLTVEGVKSVLIENTAKEVLEKQVLAAMDAPFCVLEL"
 293..435
 /gene="a1aB"

CDS	/locus_tag="SPC14_0002" /pseudo join(293..352,349..435) /gene="a1B" /locus_tag="SPC14_0002" /pseudo /codon_start=1 /transl_table=11 /product="putative oligopeptide-binding protein A1B (psedogene)" 311..379 /gene="a1B" /locus_tag="SPC14_0002" /note="1 probable transmembrane helix predicted for SPC0654 by TMHMM2.0 at aa 7-29" /pseudo complement(433..978) /gene="tnp" /locus_tag="SPC14_0003" /pseudo complement(433..978) /gene="tnp" /locus_tag="SPC14_0003" /note="member of Homology group 0004 74" /pseudo /codon_start=1 /transl_table=11 /product="putative IS630-SpiI transposase (fragment)" 1200..2645 /gene="wzg" /locus_tag="SPC14_0004" 1200..2645 /gene="wzg" /locus_tag="SPC14_0004" /note="member of Homology group 0000 90" /codon_start=1 /transl_table=11 /product="Integral membrane regulatory protein Wzg" /protein_id="CA13317.1" /db_xref="GI:68642997" /translation="MSRRPKRSQKRVSNVILLTYLLIVCFLLPIEKNTLAFRLNLVVALLVALVGLLIYKKAERFTILVFSLIVSSVLFVVOQFVLNRLNATNSYSISVAFLADSDIENTVLTGVTAPGTDMENIOKLADIKSQNDLTRVOSSSYLAAYSLIAGETKAIIVNSVFENITSESYPDYASIKIKLKGFKYKEAARKTKXQSFPVIYSGIDTGPATISSVRSDNIIMTVNRDPTKTLITTPRDAAVPVADGGNNQDKLTHAIGIYGVDSIRHTLENLYGVDIYYRNLNFTSFKMIDLIGDVANDGBESALHGKHFRPVGAVHLDSLEQLGRVEREYSLADGDNRGNQOKVIALLOKLTSTBALKNYSTIDLSIQTNMPLETMINLVNAQLSEGSNGYKVNSODLKGRTDLPSYAMPDNILYMEIDDSLAVVAKAIDVMEGR"
misc_feature	1200..1316 /gene="wzg" /locus_tag="SPC14_0004" /note="Signal peptide predicted for SPC0656 by SignalP 2.0 HMM (Signal peptide probability 0.999) with cleavage site probability 0.540 between residues 39 and 40" join(1248..1307,1335..1403,1422..1490)
misc_feature	/gene="wzg" /locus_tag="SPC14_0004" /note="3 probable transmembrane helices predicted for SPC0656 by TMHMM2.0 at aa 17-36, 46-68 and 75-97" 1407..1751 /gene="wzg" /locus_tag="SPC14_0004" /note="HMPfam hit to PFO2916, DNA polymerase processivity factor, score 5.4e-63" 1929..2372 /gene="wzg" /locus_tag="SPC14_0004" /note="HMPfam hit to PF03816, Cell envelope-related transcriptional attenuator, score 5.2e-70" 2647..3378 /locus_tag="SPC14_0005" /gene="wzh"
gene	2647..3378 /gene="wzh" /locus_tag="SPC14_0005" /EC_number="3.1.3.48" /note="member of homology group 0001 90" /codon_start=1 /transl_table=11 /product="Protein-tyrosine phosphatase wzh" /protein_id="CA133318.1" /db_xref="GI:68642998" /translation="MIDHSIVFDVDGPKSRKALLIESVRQCVRTIVTSNHRKMFETPREKIAENFILOYREIAKEVADDVIAVGAELYYTLDALEKIEKPEIPINDSKYALIEFSMTSTYRQIHFGSNIIMLGITPVTAITEYDALENNEKVARRELIDGCTTQINSHVSKRPFEKKRKFMKKRARPLYERLDVHVVASDMHNIDSRPPYMQAAVDIIAKYGAKKAELEFVNPRKIIMDQLI" 2959..3273 /gene="wzh" /locus_tag="SPC14_0005" /note="HMPfam hit to PFO2811, PHP, C-terminal, score 1.5e-10" 3387..4079 /gene="wzd" /locus_tag="SPC14_0006" 3387..4079 /gene="wzd" /locus_tag="SPC14_0006" /note="member of homology group 0002 90" /codon_start=1 /transl_table=11 /product="capsular polysaccharide biosynthesis protein wzd" /protein_id="CA133319.1" /db_xref="GI:68642999" /translation="MKRONTEIDVLOLSRALMRKXVIIIVALTSSVAFASTFVLIKPEFTSTRRIYVNRDGEKSGLTNQDLQAGSIYVKRYREIILSQDLIEVVSUKLDDLTPKGLANKIKVYLPVDTRIIVSVANRVEPEARINSIREVAOKIISTRVSDVTLLEARPALISPSPNIKRNLLIGPLAGVSGTSVIFLLEFLNTRVKRPEDIENTLOMTLIGVVPNISKUK" 3405..3821 /gene="wzd" /locus_tag="SPC14_0006" /note="HMPfam hit to PFO2706, LipoPolysaccharide biosynthesis, score 4.9e-51" join(3453..3518,3924..3992) /gene="wzd" /locus_tag="SPC14_0006" /note="2 probable transmembrane helices predicted for SPC0658 by TMHMM2.0 at aa 23-44 and 180-202" 3927..4004 /gene="wzd" /locus_tag="SPC14_0006" /note="PS00217 Sugar transport proteins signature 2." 4089..4772 /gene="wze" /locus_tag="SPC14_0007" 4089..4772 /gene="wze" /locus_tag="SPC14_0007" /note="member of homology group 0003 90" /codon_start=1 /transl_table=11 /product="tyrosine-protein kinase wze" /protein_id="CA133320.1" /db_xref="GI:68643000" /translation="MPTLEIAQQKLEFIKAEFYNALCTNIOLSGDKLKVITSVSNPGEGETTSVINIASFARAGYKTLIDGDRNSVISGVFSRKTIGLTFFLGSTADLSYGLCDTNIENTLFWOOSGVSPPNLTALOSKFNDMETLRKFDYIIVTPPGIYIDAVIIRQCADASILVTATGEANKRDVQKARQOLEQGCKLFGVIANLTDISVQKGYVASYENTGKK" 4788..6155 /gene="wcha" /locus_tag="SPC14_0008"
CDS	2647..3378 /gene="wzh" /locus_tag="SPC14_0005" /EC_number="3.1.3.48" /note="member of homology group 0001 90" /codon_start=1 /transl_table=11 /product="Protein-tyrosine phosphatase wzh" /protein_id="CA133318.1" /db_xref="GI:68642998" /translation="MIDHSIVFDVDGPKSRKALLIESVRQCVRTIVTSNHRKMFETPREKIAENFILOYREIAKEVADDVIAVGAELYYTLDALEKIEKPEIPINDSKYALIEFSMTSTYRQIHFGSNIIMLGITPVTAITEYDALENNEKVARRELIDGCTTQINSHVSKRPFEKKRKFMKKRARPLYERLDVHVVASDMHNIDSRPPYMQAAVDIIAKYGAKKAELEFVNPRKIIMDQLI" 2959..3273 /gene="wzh" /locus_tag="SPC14_0005" /note="HMPfam hit to PFO2811, PHP, C-terminal, score 1.5e-10" 3387..4079 /gene="wzd" /locus_tag="SPC14_0006" 3387..4079 /gene="wzd" /locus_tag="SPC14_0006" /note="member of homology group 0002 90" /codon_start=1 /transl_table=11 /product="capsular polysaccharide biosynthesis protein wzd" /protein_id="CA133319.1" /db_xref="GI:68642999" /translation="MKRONTEIDVLOLSRALMRKXVIIIVALTSSVAFASTFVLIKPEFTSTRRIYVNRDGEKSGLTNQDLQAGSIYVKRYREIILSQDLIEVVSUKLDDLTPKGLANKIKVYLPVDTRIIVSVANRVEPEARINSIREVAOKIISTRVSDVTLLEARPALISPSPNIKRNLLIGPLAGVSGTSVIFLLEFLNTRVKRPEDIENTLOMTLIGVVPNISKUK" 3405..3821 /gene="wzd" /locus_tag="SPC14_0006" /note="HMPfam hit to PFO2706, LipoPolysaccharide biosynthesis, score 4.9e-51" join(3453..3518,3924..3992) /gene="wzd" /locus_tag="SPC14_0006" /note="2 probable transmembrane helices predicted for SPC0658 by TMHMM2.0 at aa 23-44 and 180-202" 3927..4004 /gene="wzd" /locus_tag="SPC14_0006" /note="PS00217 Sugar transport proteins signature 2." 4089..4772 /gene="wze" /locus_tag="SPC14_0007" 4089..4772 /gene="wze" /locus_tag="SPC14_0007" /note="member of homology group 0003 90" /codon_start=1 /transl_table=11 /product="tyrosine-protein kinase wze" /protein_id="CA133320.1" /db_xref="GI:68643000" /translation="MPTLEIAQQKLEFIKAEFYNALCTNIOLSGDKLKVITSVSNPGEGETTSVINIASFARAGYKTLIDGDRNSVISGVFSRKTIGLTFFLGSTADLSYGLCDTNIENTLFWOOSGVSPPNLTALOSKFNDMETLRKFDYIIVTPPGIYIDAVIIRQCADASILVTATGEANKRDVQKARQOLEQGCKLFGVIANLTDISVQKGYVASYENTGKK" 4788..6155 /gene="wcha" /locus_tag="SPC14_0008"

CDS
4788..6155
/gene="wcha"
/locus_tag="SPC14_0008"
/note="member of homology group 0005 66"
/codon_table=11
/product="undecaprenylphosphate glucosylphosphotransferase
wcha (initial sugar transferase)"
Query Match 12.7%; Score 888.8; DB 1; Length 19918;
Best Local Similarity 62.8%; Pred. No. 6,4e-81;
Matches 1484; Conservative 0; Mismatches 857; Indels 21; Gaps 6;
QY 16 TGGCATATTGATATGATAGCAGTTGCAATTCCTGCAATCTTAACAAGCATATACCA 75
DB 4816 TGGCATATTGATAGCATATATTTGATTTATGTTATTTCTTAGCTTTGTTAGAG 4875
QY 76 ATGCTGATTTAAATCGTTCGAAATTTTATTCATTAATGATGTTCAATTTTGGCAATTTT 135
DB 4876 AGACAGAACTGAAACGTTCTCGATGTTATACCTATACCTTCACACTTTTGTATTCCT 4935
QY 136 TTATATCTCGTATGCCAGTTGAATTGAGTATAGGTAATCGATAGAGTTGAAAAA 195
DB 4936 ATTTTAGTTCCTATGTTAAACAAATTTTTTAAAGAGGTAACATGAGTTAAATGTA 4995
QY 196 CATTACTATGATATATATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGGAGA 255
DB 4996 CTATAGATATATTTTCTTTGCAATAGCTATAGTATTAATCTTTTATATAGCG 5055
QY 256 ATAAATTCGCACCTTTCAGACGTGTCGCGCTGATTTTACATTATTAATCTGCTTTGG 315
DB 5056 AAGGTTTAGTATCTCTAGAAAGAAATGTTATACCTTAACTTGAAGAAATACCT 5115
QY 316 TATACCTATTTAAAGTATTTAAGCAGTTTAAGATACCTTTCTATTTTCCACATCT 375
DB 5116 TATACCTGTTAAATTTCTTGTAAAGAAATATGGAAGCATGTGTTTTTAATCCAAAA 5175
QY 376 ATCAAAAAAGAGATCTTAATTACACGCGTGAACGATGGAAAAATATGCAAGTTTAT 435
DB 5176 ATAGCAAAAAATTTTACGTGTAAACAGTAACGAAAAATAGAAAAAGTTCTTGATTAAT 5235
QY 436 TTGAATCACATTAACAATTCAAAAAATCTGTTGCAATGTTAGTTTAAAGTACAGAA 495
DB 5236 TGCTAGAAATCTGAATGAACCTTCAAGGAACTGATGAGTAAGTGTGTTGATTAATCTG 5295
QY 496 TAGATTAATTAATTTATCATTAACCGCTCTATTTTCTGAGAAAGCTATAGAGTTT 555
DB 5296 ATTTTCAACATGAATAAATACCTGTAATGAAA-----GAAAAAATTAATGAATTTG 5349
QY 556 CAACAAGGAAGTGTGACACAGCTCTTAATTAATCTAACCAAGTGAATTTTGAACGTA 615
DB 5350 CAAGCATGAAGTGTGATGAGGTGTTGTCGATCTCCAGGAGAAAGCTATGATTTG 5409
QY 616 AGCAATTCGTTTCAGATTTTGAAGTTGTTAGGTATGTAAGCGGTGATATTAATTCAT 675
DB 5410 GAGAAATTAATCTAAAGTTTGAACATGGGAGTAAATGTAACGTAAATCTTAATCAT 5469
QY 676 TCGGTTTACTGCGTGAACAAAAAATAATCAACTGCTAGGTGACCATAGCAATTTGTA 735
DB 5470 TTAATAAGAAATTTGGTCTGAATTAACAATTCATGAGATGAGTATGTAATGATGCA 5529
QY 736 CTTTTCACAAATTTTAAAGCTAGCTATATCAGATGAAACGACTTTGATATAC 795
DB 5530 CTTTCTCTACAAATTTTAAATTAACCTAGTCAATGATTTCAAAAGAAATTCAGATTTT 5589
QY 796 TCGAGCGGTAGTGGGTTAATTAATTTGTTGTTAGTTTCTAATTTTGTAGTTCCAAATTA 855
DB 5590 GTGATGCACATATGGCTTATCTTTTGTCTATAGCTATAGCTTATATTCATTTGA 5649
QY 856 TTGCTAGAGATGTGAGACCGGCTATTTTGTCTCAGAAAAGATTGAGACAGATGAGCGCA 915
DB 5650 TTGCTAAAGATGGGAGCACAGCTATTTTGTCTCAAACTCTATATAGGAAAAATGATGTCAC 5709

QY 916 TATTTACATTTCTACAAAGTTTGCATGATGATATGTTGATCTGAGAGGCGCAAAAAAGACT 975
DB 5710 ATTTTACCTTTTATTAATTCGTTGATGCGGATGAGTGAATTTATTCAAAGAACACT 5769
QY 976 TGCTGAGCAAAAACAGATGCAAGGTTGGTATGTTTAAATGGAAGAAACGATCTAG 1035
DB 5770 TGATGATCAAAATACGATGCGAGGTGATCTTT-----AAGATGCAATATGATCTCG 5824
QY 1036 AATTACTCCAAATTTGACATTTTCATACGCAAAAAAGTTTGAACGATTTACCAAGTTT 1095
DB 5825 TGTTCAAAATAATGATCGCTTATTCG--TAAACAGTTTATAGATGATTAACCCAGTTT 5883
QY 1096 ATAAATGTTTAAATGCGCATATGATCTATAGTTGTATACAGTCCACATCAAGTTATGAT 1155
DB 5884 GGAATGCTTTTATAGAGATATGAGTATGTTGGTGGAAACGTTCCACTCACTAGTACAGAT 5943
QY 1156 TTGAAAAATATCTCCTGCTCAAAAAGACATGATGTTTAAACAGGATTTAAGATC 1215
DB 5944 ATGATCAGTATCTCCAGAACAGAAACGTGCACTCAGCTTTAACTGCTATTAACAGTT 6003
QY 1216 TCTGCAAGTTAGTGTGCTAGTAAATATCAAGACTTCGACGACGATGTTGCTGACT 1275
DB 6004 TATGCAAGTTAGCGGCGGTATGTAATAATACCATTTGACGATGTTGTAATTAAGAG 6063
QY 1276 TAGCATCACTGATTAATGGAATATCTGCTCAGATTTTAAATTTTATTAAGACAGTGA 1335
DB 6064 TGGCTTATATGATTAATGGAACATCTGAAAGATATGAAATTTGCTTAAACCTGTA 6123
QY 1336 AAGTGTATGTTGAGAGAGGGAAGTATGAAAGTATGAAAGTTTGTGTCGCTT 1395
DB 6124 AAGTGTATTTATGAAATAATGAGCAAGTAAAGG---ATGAAGATATCTAGTTGAT 6180
QY 1396 CTTCAAGGGGACATTTGACTCACTGATTTGTTAAACCGTTTGAAGAAAGAAAC 1455
DB 6181 CTAGTGAAGTCAATTTAATCTCATTTGTTATTTAAACCTTCTGGAAGATAGAGAA 6240
QY 1456 GTTTTGGGTAACTTTGATTAAGAGAGATGCAAGAACTTTTGAAGAAATGAAAAATCT 1515
DB 6241 GATTTGGGTGACTTTGATTAAGAAAGATGCAAGAAATTTTGAAGAAATATTTT 6300
QY 1516 ATCCATGTTACTTCCAGAAATCGCAATCTCATTAATTTAGTGAATAATCTTCTTAG 1575
DB 6301 ATCCATGTTATTAATCTTACTAATCTTAATCTTAATAATCTTATTAATAATFACATCTTG 6360
QY 1576 CTTTCAAAATTTTACCTGATGAGAAACAGATGTTATTAATTCATCTGTCGCGCGTTG 1635
DB 6361 CTTTCAAAATTTTGAAGAAAGAACGCCGTGATTTATTCGCTCATCAGGGGACGCTGAG 6420
QY 1636 CTGTCCCTCTTTTACATCGGAAACATTTTGGAGCAAAAGCATTTATTTAGAGTAT 1695
DB 6421 CAGTTCCTTCTTTATCTTAGGAAATATTTGGTCTTAAGACAGCTATATATGAAATGAT 6480
QY 1696 TTGATGAGTTAATTAATCTACATTAATCTGAAATCAGTTTATCCGTAACGATATTT 1755
DB 6481 TTGATGAATTAATGCTCCGACATGACTGGAACTGTTTATCCAGTATCTGACAGGT 6540
QY 1756 TTAATGTTAGTGGAGAAATGAAGAGTATATCTAATCTAATTAATCTTGGGAGTA 1815
DB 6541 TTAATGTAATGAGAAAGAAATGAAGAAATATATCCGAAAGCAATCAATCTGGAGGA 6600
QY 1816 TTTTAAATGATTTTGTAAACAGTAGAACCTCATGAACACAGTTTATGATTAATA 1875
DB 6601 TTTTCTAATGATTTTGTGACAGTAGTACTCATGAACACAGTTTATGACTTTATTA 6660
QY 1876 AGGATGATTTATGAAAAAATAAGAAATATGAAATATTAACGAGAAATATTTATCAACAGG 1935
DB 6661 AGAGGTGATGATTAATAAGGTGAGGATTTATCAGAGATGAGTTTATTAACAAACAGG 6720
QY 1936 ATATTCGACTATATTCAGAAATATGCAAGTATTAATAAATTTCTCAGTTCAAAAGAAAT 1995
DB 6721 ATACTCTAATTAATGTTCCAAATTTTGTATATGGAAGAAATTAATCTTATGAAAAAAT 6780
QY 1996 GGAACATATATTAACAATAACAGAAATGATTAATTTGTCACGAGAGCCCGCTACTTTAT 2055

Db	6781	GAATCAATGATTTAAAGAAATCAGATATTATCTTTACCATGGCGGTCCAGCTACCTTTAT	6844
Qy	2056	GAATTCATTTATCCAAAGGAAAAAAACAATTATGTTTCTGACACAAAAAAGTATGTGA	2115
Db	6841	GGCAGTTATGCTCTAAAGCTAAAAAATCCATATATTGTTCCGGCGCTAAAAAATTTGGTGA	6900
Qy	2116	ACATGTAAATGATCATCAAGTAGTGGATTGTGAGAAATTTTACACAGTAATTAATATTTT	2175
Db	6901	GCATGTAAATGATCACCGACGATGCATTTGTAAAAATATACGAAGAAATATTCATTTAT	6960
Qy	2176	ATTATATGAAAATATAGATGATTGTTGTGAAAAAATTATGAGTT---TCTAACCAAC	2232
Db	6961	AGTATATGATGATATTTTCAGACTTACATTTAATCTTCATATTTTAAAGACAAACATTT	7020
Qy	2233	TAACTTATACATCAATATATATTTTTTTTGGAAAAATTA---ACAAATATGTGGAAA	2288
Db	7021	TGAACCTATTTTGAATTAACGAGACTTTTAATGTAAGTTCATGTGAAAAACAGTAACCT	7080
Qy	2290	ATTATATGAGGATCAAGAAATGGAATATAAAAAAGATGCATTTGTATATGCGTTAC	2349
Db	7081	TTTTTAAGGAATAATAAATAAATGAAAATTAGAATGAACTCAGATATTTCTTATATAAT	7140
Qy	2350	ATAATTTTCTCAGATTTTACT	2371
Db	7141	ACCTTGGTTATTTATCTTACT	7162

RESULT 9				
SPCPS14E				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED		</		

REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 1821)
Kolman, M.A.B.
Direct Submission
Submitted (06-SEP-1995) M.A.B. Kolman, Institute of Infectious
Diseases and, Immunology, Department of Bacteriology, School of
Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL-
3508 TD Utrecht, NETHERLANDS
revised by [4]

REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL

6
Kolman, M.A.B.
Direct Submission
Submitted (23-APR-1996) M.A.B. Kolman, Institute of Infectious
Diseases and, Immunology, Department of Bacteriology, School of
Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL-
3508 TD Utrecht, NETHERLANDS
revised by [5]

7 (bases 1 to 14943)
Kolman, M.A.B.
Direct Submission
Submitted (16-JUN-1997) M.A.B. Kolman, Institute of Infectious
Diseases and, Immunology, Department of Bacteriology, School of
Veterinary Medicine, University of Utrecht, P.O. Box 80.165, NL-
3508 TD Utrecht, NETHERLANDS
On Jun 17, 1997 this sequence version replaced gi:1279732.
Location/Qualifiers
1. .14943

```

gene
    <1..353
    /gene="dexB"
    <1..353
    /gene="dexB"
    /function="degradation of glucans"
    /codon_start=3
    /transl_table=11
    /product="alpha, 1-6-glucosidase"
    /protein_id="CA59784.1"
    /db_xref="GI:2198540"
    /db_xref="GOA:Q54796"
    /db_xref="UniProt/Swiss-Prot:Q54796"
    /translation="EWLVAVNPYEMINVOBALNPISFYTKLVQIRKNSWLVA
DPEILDTRDKYPAIRKDGRRFLVYANLSESDLTVEGKYSVLIENTAAEYLER
QVLAWPAFCVELL"
    700..753
    join(1054..10552,10568..11416)
    /gene="cps14K"
    1054..10552
    /gene="cps14K"
    1196..1305
    /gene="cps14K"
    1360..1365
    /gene="cps14K"
    1383..1388
    /gene="cps14K"
    1408..1410
    /gene="cps14K"
    /gene="cps14A"
    1408..1410
    /gene="cps14A"
    1419..2864
    /gene="cps14A"
    1419..2864
    /gene="cps14A"
    /function="putative regulatory"
    /codon_start=1
    /transl_table=11
    /product="capsular polysaccharide synthesis protein"
    /protein_id="CA59785.1"
    /db_xref="GI:2198541"
    /db_xref="GOA:O07338"

```

gene /db_xref="InterPro:IPR004190"
/db_xref="InterPro:IPR004474"
/db_xref="UniProt/TREMBL:O07338"
/translation="MSRRFFKRSQKVRSVNIVLTITLYLLVCGELLFLIKNTIAE
RLNIVLVATLVVALVGLILLIYKKAERFIFLIVSILVSVLFAVQOPGLNTR
LNATNSYSGSVLAVLADSDIENYVLTSTAPGSDENIIOKLADISQNTDIT
VNOSSYLAAYSKLIGETKRAIVNSVFNENIIESEYDPYAKIKKIYTKPTKVAEP
KTSKQSGFNITYSGIDTYGPIISSVSRSRVNITLMTNPDTKILLTTTPRAYVYDQ
GNNQDQKTHAIGYDVSSITLLENLYGVADINYYVRLNFISFLMKIDLLGVDVHNDQ
EESALHGKHFPGVNGHLDSEQALGFVERYSIADGDRGRNQOKVI VAILQKLTST
EALKVNSTIIDLQDSIOTNNPLEMTMINLVNTQLESGGNYKVNQDLKGTGRDLP
AMPDNIYVMEIDSSIAVVKAIQDVMEGR"

gene 2852. .2857
/gene="cpel14B"
RBS 2852. .2857
/gene="cpel14B"
gene 2866. .3597
/gene="cpel14B"
CDS 2866. .3597
/gene="cpel14B"
/codon_start=1
/transl_table=11
/product="capsular polysaccharide synthesis protein"
/protein_id="CAAS9774.1"
/db_xref="GI:1279733"
/db_xref="GOA:P72510"
/db_xref="InterPro:IPR004013"
/db_xref="UniProt/TREMBL:P72510"
/translation="MDIHSHVFPVDPOPKRSERSKALLIESYRQGVRTIVSTSHRR
KMFERPERKIAENFLQREIATAKEVADLVIAAGAEIYITLDALKEKKIIPLNDS
RIALIEFSNTHSRQIHGSLNITLMGITTPTVAHIERDALENNEKHYREIDMGCTT
QINSHVSYSPKFEFGKFKPKRRARYFERLDVHVASDMHNDSPRYMOAADITIA
KRWSSKSERTEFVDNPKRIMQDLI"

gene 3596. .3600
/gene="cpel14C"
RBS 3596. .3600
/gene="cpel14C"
gene 3606. .4298
/gene="cpel14C"
CDS 3606. .4298
/gene="cpel14C"
/function="putative chain length regulator"
/codon_start=1
/transl_table=11
/product="capsular polysaccharide synthesis protein"
/protein_id="CAAS9775.1"
/db_xref="GI:1279734"
/db_xref="GOA:P72511"
/db_xref="InterPro:IPR003856"
/db_xref="InterPro:IPR005701"
/db_xref="InterPro:IPR005829"
/db_xref="UniProt/TREMBL:P72511"
/translation="MKEONTLSDIYGLSRALMKRYLILVAITSSVAFVSFEVLA
KPEFISTIRIYVNRDQEGSKLTNODLQAGYILVKDREITLSGYULEEVSVDKLHL
LTPKGLAMIKIVPVDTRIIVSVNDRVPEASRIANSLREVAOKIISTIRSDVAV
TLEAPRALSPSSPNIKRNTLIGFLAGVSGTSIVFLFELNTRVKRPEDIENTLQMTM
LIVGVNLSGLK"

gene 4297. .4302
/gene="cpel14D"
RBS 4297. .4302
/gene="cpel14D"
gene 4308. .4991
/gene="cpel14D"
CDS 4308. .4991
/gene="cpel14D"
/function="putative polysaccharide export protein"
/codon_start=1
/transl_table=11
/product="capsular polysaccharide synthesis protein"
/protein_id="CAAS9776.1"
/db_xref="GI:1279735"
/db_xref="GOA:P72512"
/db_xref="InterPro:IPR005702"

		/db_xref="UniProt/TREMBL:P72512"		
		/translation="MPTLEIAQKYLEFKIKAEYVNAALCTNQLSGDKLKVISVSN		
		PBEGRKTTSVNIAISFARQAQKLIDGDFRNSVIGSFRKRITGLIEFLPSGTADL		
		SVGLCDTNIENLFVQSSEVPNTALLOSKEFNMDIETLRKFVDIVDPVGVIAT		
		DAVILTRCKDCASIIIVATGEANRDVQAKQOLETGKFLGVVLNKLDISVKGYV		
		AAGYNNGKK"		
gene		4995..5000		
		/gene="cps14E"		
RBS		4995..5000		
		/gene="cps14E"		
gene		5007..6374		
Query Match	12.7%;	Score 885.6;	DB 1;	Length 14943;
Best Local Similarity	62.7%;	Pred. No. 1.5e-80;	Mismatches 859;	Indels 21; Gaps 6;
Matches 1482;	Conservative 0;			
Dy	16	TGGCATTATTGGATGATGATGACGAACTTCGCAATCTTTAACAGTCATATACC	75	
Dy	5035	TGGCGATTATTCAGAGTATTATGTCTATTATTTGGTTTATTTCTTAGCTTGTTAG	5094	
Oy	76	ATGCTGATTTAAATCGTTCTGGAATTTTTATCATATGATGATGTTCAATTTTG	135	
Dy	5095	AGACAGAATTTGAACGTTCTGCATGGTATACTAATCTTCACCTTTTGTATTCT	5154	
Oy	136	TTATATCTCGTAAGCCAGTTGAAATTTAGATATAGGTATATCGATAGAGTTGA	195	
Dy	5155	ATTTTAGTTCCTATGTTAACAAATTTTTTAAAGAGGTACCTAGTGAATTATAT	5214	
Oy	196	CATTATACATATGATAATATATTTGCAATTTTCTTAGCGCAGTATCAATTTTGT	255	
Dy	5215	CTATAGATATATTTTCTTTCGCCATTAAGTATATGATATTAACCTTTTATAG	5274	
Oy	256	ATATATTCGACCTTCAAGCGTGGCCGGTATTTCACTTAATATATTAACCTTGG	315	
Dy	5275	AACGTTTAGATCTAGAAAGAAATGGATGATCTTTAACTTTAAGAAATATCT	5334	
Oy	316	TATACCTATTTAAGTAATTTATTAAGACATTTAAGATAGCTTCTATTTTGC	375	
Dy	5335	TATACCTGTTAAATTTCTTAGTAAAGAAATATTTGAAGCATGTGTTTTTATC	5394	
Oy	376	ATCAAATAAAGACATCTCTAATTACAACGCGTGAACATGGGAATAATGCAAT	435	
Dy	5395	ATAGCAAAAAATTTTACTGTAAACATTAACGAAAAATATGAAAAGTCTTAT	5454	
Oy	436	TTGATCACATAAACAAATTCAAAAAATCTTGTGCAATGGTAGTTTAGTAC	495	
Dy	5455	TGCTAGATCTGATGAACCTTACATGGAACCTGGTAGAGTAAAGTTTGGAT	5514	
Oy	496	TAGATAAATTAATTTATCATTAACCGCTCTATTTCTGTGGAAGAGCTATAG	555	
Dy	5515	ATTTTCAACATGATAAATATCTGTAAATGAAA-----GAAAAATTAATGA	5566	
Oy	556	CAACAAGGAAGTGTGACACGCTCTTATTAATCTTACAACAGTGAATTTTGA	615	
Dy	5569	CAAGCATGAATGTGTGATGAGTGTGTGCAATCTTCCAGAGAGAGCTATGAT	5628	
Oy	616	AGCAATTCGTTTCAATTTTGATGTGTGATGATGATGATGATGATGATGAT	675	
Dy	5629	GAGAAATTAATCTAAGTTTGAACAATGGGATAGATGTAATCTTAATCAT	5688	
Oy	676	TCGGTTTACTGCGTTGAAAAACAAAAATCCAACCTGCTAGTGAACATAG	735	
Dy	5689	TTAATTAAGAAATTTGGTCTGTAATTAACAAATTCATGATATGATGATGAT	5748	
Oy	736	CTTTTTCACAAATTTTATTAAGCTAGTCAATCATGATGATAAAGCACTTTG	795	
Dy	5749	CTTTCTCAACAAATTTTATTAACCTAGTCAATGATTTCAAAGAGATTTCT	5800	
Oy	796	TCGAGCGTATGTCGGGTAAATTAATTTGTGATATAGTTTCTAATTTTGT	855	
Dy	5809	GTGTGTCACATATGTGCTTATTTCTTTTGGATATGATGATGATTTTAA	5866	
Oy	856	TTCTGATGATGTGTGACCGGCTATTTTGTCTCAGAAACGATTTGACATG	915	

Db 5869 TTGTAAGAGGCGGACGAGTATTTTCTCAAACTGATATAGGAAAAATGCTGAC 5928
Qy 916 TATTTACATTCACAAATTTCATTCATGATGATGCTGAGGAGCGGAAAAAACT 975
Db 5929 ATTTTACCTTTTAAATTCCTGATGCGGATGCGATGCTGAAGTTATCAAGAACAT 5988
Qy 976 TGTCTGCGCAAAACAGATGCAAGGAGGTGATGTTTAAATGGGAAAAACATCTAG 1035
Db 5989 TGTATGATCAAAAATGATGCGAGGTGATGTTT-----AAGATGACATATCTCG 6043
Qy 1036 AATTAATCCAAATTGACATTTCAACGCAAAAACAAATTGACAGATTACCAATTTT 1095
Db 6044 TGTACCAAAAATGATGCTTTATGCG- TAAACCAAGTTAGATGAGTTACCCAGTTT 6102
Qy 1096 ATATATGTTTAAATGGCGATGATGATGCTAGTGTACAGCTCCACTACGATTGAT 1155
Db 6103 GGAATGCTTTTATAGAGATGATGATGTTGGTGGAAACAGTCCACCTACAGTACAG 6162
Qy 1156 TTGAAAAATATATCTCGTCAAAAAGAGAGATTGATTTTAAACGAGGATTACAGCT 1215
Db 6163 ATATGATGATATCTCAAGAACAGAAACGTCGACTCAGCTTTAACTGATATACAGTT 6222
Qy 1216 TCTGCGAGTTAGTGTGCTAGTAAATATCAAGACTTCGACGAGTGTGCTGCTGACT 1275
Db 6223 TATGCGAGTTAGCGGCGCTAGTAAATACGATTTTGAAGATGTTGTAATTTAGAT 6282
Qy 1276 TACGATACATTTAAATTTGATGCTATCTGCTAGATATTAATTTTAAAGACATGA 1335
Db 6283 TGGCTTATATTTGAATATTTGACAACTTGAAAGATATTTGCTTAAACCTGTTA 6342
Qy 1336 AAGTTGATTTGTTGAGAGAGGAAATGATTAAGAAAGTTTGTGTGCTGCTT 1395
Db 6343 AAGTTGATTTTGAAGAAATGAGAGCGAATGAAAGG--ATGAAAGATATGCTAGTGGAT 6399
Qy 1396 CTTCAGAGGAGACATTTGACTCATCTGTTTGTAAAAACGTTTGGAGAGAAAGAAC 1455
Db 6400 CTAGTGAAGTCACTTACTCATTTGTATTTTAAACCTTTCTGGAAGATAGAGAA 6459
Qy 1456 GTTTTGGGTAACTTTGATTAAGAGAGATCAAGAGCTTTTGAAGATGAAGAAATGT 1515
Db 6460 GATTTGGGTGACTTTTGTATTAAGAAAGATCAAGAGATTTTGAAGAAATGAAATATTT 6519
Qy 1516 ATTCATGTTACTTCCAAACATTCGCAATCTCATTAATTTAGGAAAAATCTTTCTAG 1575
Db 6520 ATCCATGTTATTTCTCACTCAATCTGATATTAAGAAATCTTATTAAGAAATCTTT 6579
Qy 1576 CTTCAAAAATTTTACGATGAGAGAACAGATGTTATTTTCACTGCTGCGGCGCTTG 1635
Db 6580 CTTCAAAAATTTTGAAGAGAGAACGCTTATATTTCTGCTATCAGGCGGCGCTGAG 6639
Qy 1636 CTGTCCCTCTCTTTTACATGAGAAACATATTTGGAGCAAGCAATTTATTTGAATAT 1695
Db 6640 CAGTTCCTTTCTTTTATCTAGGAAATTTTGGTGTAGACAGTCTATATATGAAATAT 6699
Qy 1696 TTGATGAGTTATTAATCTACATTAACGAAAACTAGTTTATCCGTTACAGATTTT 1755
Db 6700 TTATATGAATTTGATGCTCCACATGATGCGAAACCTGTTTATCCAGTACGACGCT 6759
Qy 1756 TTATGTTGCTGAGGAGAAAGATGAAGAGATATCTTAATCTATTTAATCTTGGGAGTA 1815
Db 6760 TTATGTTGCTGAGGAGAGATGAAGAGATATCTTAATCTTGAATCTCGAGAGGA 6819
Qy 1816 TTTTAAATGATTTTGTAAACAGTGAACATCAAGCAAGTTTATGATGATTAAT 1875
Db 6820 TTTTCTATGATTTTGTGACAGTATGCTCATGAAACAGTTTATGATGATTAAT 6879
Qy 1876 AGGATGATTTTATGAAAAAATGAAATGATTAACGAGAAATATTTTCAAAACAG 1935
Db 6880 AGAGGTTGATGATTAAGAGTGAAGATTTTATGAGATGATTTTATCAAAACAG 6939
Qy 1936 ATATTTCTGATATATTTCAAGATATTTCAAGTATTAAGAAATTTCTCAGTTCAAGAAAT 1995

Db 6940 ATACTGAATATATGTTCCAAAATTTTGTAAATGGAAAAATTAATATCTTATGAAAAAT 6999
Qy 1996 GGAAACATATATTAACAAATCAGAAATGATTTATTTCCACGAGGCGCTACTTTAT 2055
Db 7000 GAATCAATTTGATTAAGAAATCAGATTTATCATTTACCCATGGGCTCAGCTACGTTAT 7059
Qy 2056 GAATTCATTTATCCAAAGAAAAAACAATTTATTTTCTTACGAAAAAAGATGCTGA 2115
Db 7060 GCGATTTATTTGCTTAAAGATTAAGAAATCAATTAATTTGTTCCGCGCTAAAAAATTTGCTGA 7119
Qy 2116 ACATGTAATGATCATCAAGTATGATTTGTAAGAAATTTTACAGATTAATATTTT 2175
Db 7120 GCATGTAATGATCATCAAGTATGATTTGTAAGAAATTTTACAGATTAATATTTT 7179
Qy 2176 ATTTATGAAAAATATGATGATTTGTTGAAAAAATTTATGAGTT---TCTAAGCAAC 2232
Db 7180 AGTTATGATGATATTTGATGATTTGATTTATTTATTTTATTTTATGAGCAAAACATTT 7239
Qy 2233 TAACTTACATCAATTAATTAATTTTGTGTAAGATTA---ACAAATAGTTGAAA 2289
Db 7240 TGAATCTATTTGATTAAGAGATTTATGATGATTTCAATGTTGAAATGATGATCT 7299
Qy 2290 ATTTAATGAGATCAAGAAATGATTAATTAAGAAATGATGATTTGATGATGCTTATC 2349
Db 7300 TTTTAAAGAAATTAATTAATTAATGAAATTAAGATGAACTCAGTATTTTCTTATTAAT 7359
Qy 2350 ATTAATTTTCTCAGATTTACT 2371
Db 7360 ACCTTGTGTTTATTAATCTACT 7381

RESULT 10
CR931666
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 22405
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="688/6"
/db_xref="taxon:1313"
<1..134
/gene="dexB"
/locus_tag="SPC15F_0001"
<1..134
/gene="dexB"
/locus_tag="SPC15F_0001"
/codon_start=3

Streptococcus pneumoniae
Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 22405)
Bentley, S.D., Aanensen, D., Mayroldi, A., Saunders, D.,
Rabinowitz, E., Collins, M., Donaghy, K., Harris, D., Kalfof, M.S.,
Murphy, U., Quail, M.A., Samuel, G., Skovsted, I.C., Battey, B.G.,
Reeves, P., Parkhill, J., and Spratt, B.G.
Genetic analysis of the capsular biosynthetic locus from all 90
serotypes of Streptococcus pneumoniae
Unpublished
2 (bases 1 to 22405)
Bentley, S.D.
Direct Submissions
Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: sdb@anger.ac.uk
NOTE: This sequence was generated from a PCR product representing
the region from dexB to a1a and is not necessarily responsible for
the expressed capsule serotype. For a detailed description of how
CDS products were predicted see the associated publication.
Location/Qualifiers


```
gene
    /transl_table=11
    /protein_id="CAI33402.1"
    /db_xref="GI:6864309"
    /translation="SNEEQDLVEGKVKSVLENTAKKEVLEKQVLAPMDARCVELL"
    293..466
    /gene="a11b"
    /locus_tag="SPC15F_0002"
    /pseudo
    293..466
    /gene="a11b"
    /locus_tag="SPC15F_0002"
    /pseudo
    /codon_start=1
    /transl_table=11
    /product="putative oligopeptide-binding protein A11B
    (pseudogene)"
    /db_xref="PSEUDO:CAI33403.1"
    311..364
    /gene="a11b"
    /locus_tag="SPC15F_0002"
    /note="1 probable transmembrane helix predicted for
    SPC0740 by TMHMM2.0 at aa 7-24"
    /pseudo
    complement(join(459..1036,1038..1322))
    /gene="tnp"
    /locus_tag="SPC15F_0003"
    /pseudo
    complement(join(459..605,605..1036,1038..1322))
    /gene="tnp"
    /locus_tag="SPC15F_0003"
    /note="member of homology group 0004 74"
    /pseudo
    /codon_start=1
    /transl_table=11
    /product="putative IS630-Spn1 transposase"
    complement(933..1304)
    /gene="tnp"
    /locus_tag="SPC15F_0003"
    /note="HMMpfam hit to PF01710, Transposase, Synechocystis
    PC 6803, score 9.1e-11"
    /pseudo
    1544..2998
    /gene="wzg"
    /locus_tag="SPC15F_0004"
    1544..2998
    /gene="wzg"
    /locus_tag="SPC15F_0004"
    /note="member of homology group 0000 90"
    /codon_start=1
    /transl_table=11
    /product="integral membrane regulatory protein Wzg"
    /protein_id="CAI33405.1"
    /db_xref="GI:68643100"
    /translation="MLMSRRFKKSNQKRSVNTVLTLYLLVCFLLFLFKNTI
    LARVNLVVTALVLVALVGLLLIYKAKEKTFLLVPSILVSSVSAPAVQFVGL
    TNLNTSNSEYSISVAVLADSDIEVNTQLTSVTPGTGDNENIOKLADISQNT
    DLVVDSSSYLAAYKSLIAGETKAIIVNSVEVNIISEYDASKIKIYTKGFTKV
    EAKTSKNSQFNIVSGIDITGPRISSVSRDVAIIMTVNMDTKILLTTPRAYPI
    ADGQNNOKDLTHAGIYGVDSIHTLENLVGVINITYRLNFTSPFKMIDLLGVVH
    NDEFTLHGKFFPVGNVHLDSEQALGFVRERYSLADGRDRGRNQQYVAILIKL
    TSTEVKNYSIILOGLQDSIQTMPIETMIDLNTOLSESGNYKNSQDLKGRMDL
    PSYAMPDSNLYWEIDSSILA VYKAAIQDWEGR"
    1544..1669
    /gene="wzg"
    /locus_tag="SPC15F_0004"
    /note="Signal peptide predicted for SPC0742 by SignalP 2.0
    HMM (Signal peptide probability 0.999) with cleavage site
    probability 0.540 between residues 42 and 43"
    join(1601..1660,1688..1756,1775..1843)
    /gene="wzg"
    /locus_tag="SPC15F_0004"
    /note="3 probable transmembrane helices predicted for
    SPC0742 by TMHMM2.0 at aa 20-39, 49-71 and 78-100"

misc_feature
    1760..2104
    /gene="wzg"
    /locus_tag="SPC15F_0004"
    /note="HMMpfam hit to PF02916, DNA polymerase processivity
    factor, Score 5.4e-63"
    2282..2725
    /gene="wzg"
    /locus_tag="SPC15F_0004"
    /note="HMMpfam hit to PF03816, Cell envelope-related
    transcriptional attenuator, score 5.9e-70"
    3000..3731
    /gene="wzh"
    /locus_tag="SPC15F_0005"
    3000..3731
    /gene="wzh"
    /locus_tag="SPC15F_0005"
    /EC_number="3.1.3.48"
    /note="member of homology group 0001 90"
    /codon_start=1
    /transl_table=11
    /product="protein-tyrosine phosphatase Wzh"
    /protein_id="CAI33406.1"
    /db_xref="GI:68643101"
    /translation="MIDIHSHTVFPVNDGPKRSRKSALLAEAYRGVITVTSRHR
    KCMETPRBKATENFLQYREIAKEVASDLVIAYGAEIYTPDVLKDKKRIPTLND
    RYALIEFSMTNPYRDHSALNKLIMLGITPVIAHIERYDALENNEKRVRELIDMCYT
    QVNSHVLVSKLFGERYKRMKRAQYFLEQDLVIAHVDMLNLDGRPPHMAEAYDLVT
    OKYGEAKQDELVDNPRKIIMDQI"
    3312..3626
    /gene="wzh"
    /locus_tag="SPC15F_0005"
    /note="HMMpfam hit to PF02811, PHP, C-terminal, score
    1.3e-12"
    3740..4432
    /gene="wzd"
    /locus_tag="SPC15F_0006"
    3740..4432
    /gene="wzd"
    /locus_tag="SPC15F_0006"
    /note="member of homology group 0002 90"
    /codon_start=1
    /transl_table=11
    /product="capsular polysaccharide biosynthesis protein
    Wzd"
    /protein_id="CAI33407.1"
    /db_xref="GI:68643102"
    /translation="MEBQNTLEIDVLQLEFRSLMKRKLVLVAITTSVAFAYSFVI
    KRFSTRIYVNRNDQEGKSLTNODLQAGTYLKVDRYEITLSDVLEEVYSDKLD
    LTPKGLANKIKYTVPYDTRIVSISYKDQPEASRIANSIREVAEKIYAVYRVSQV
    TLEEARPATTPSSPNVRNTLFGFAGAVVIAVLLLELDTRVRRPVEDVLIQIP
    LGLVPDLDMK"
    3758..4174
    /gene="wzd"
    /locus_tag="SPC15F_0006"
    /note="HMMpfam hit to PF02706, Lipopolysaccharide
    biosynthesis, score 1.6e-52"
    join(3806..3874,4274..4342)
    /gene="wzd"
    /locus_tag="SPC15F_0006"
    /note="2 probable transmembrane helices predicted for
    SPC0744 by TMHMM2.0 at aa 23-45 and 179-201"
    4280..4357
    /gene="wzd"
    /locus_tag="SPC15F_0006"
    /note="PS00217 Sugar transport proteins signature 2."
    4442..5131
    /gene="wze"
    /locus_tag="SPC15F_0007"
    4442..5131
    /gene="wze"
    /locus_tag="SPC15F_0007"
    /EC_number="2.7.1.112"
    /note="member of homology group 0003 90"
```



```

/codon_start=1
/transl_table=11
/product="tyrosine-protein kinase Mze"
/db_xref="GI:68643103"
/translation="MPTLEISQAKLDFVKAEYVNALCTNQLQSDGLKVFSTSVK
LGEKSTSTENIMAFAPARAGYKTLIDGRNSVMGVFPAKRIKGLFEFSGTDL
SQGLCTNENLFIQAGSVSPNPTALQSNFSTMLRKYEDYIYDTAEVAVI
DAIITRKCDASILVTEAGINRDIQKAEQLEHNGKPLGVLLKFPDSVVKQSY
GNYDYGKXKK"
gene
5146..6513
/locus_tag="SPC15F_0008"
CDS
12.4; Score 869.4; DB 1; Length 22405;
Best Local Similarity 62.0%; Pred. No. 5.3e-79;
Matches 1483; Conservative 0; Mismatches 886; Indels 22; Gaps 6;

Query Match
14 ATTGGCATTTTGTATGTATGATGCAATTTTCGCAATCTTAAAGTCATATACC 73
5172 ATTGGCATATATCAGAGTTTCTGTATTTATGACTTATCTACTTAGTGTGAG 5231
74 AAATGCTGATTTAAATCGTCTGGAATTTTATCATATGATGCTTATTTTGCAAT 133
5232 AGAAGGGAGATTTGTTCAACAACGCTATTGCACTTATATCCTCATTTTGTCTT 5291
134 TTTTATATCTCGATGCGCAAGTTGATTTGAGTATAGGTAACTGATAGATTGAAA 193
5292 TTATATCAGATTAATGAGCAGATTTCTTAAAGAGATATTTGATTTGAACTTGCA 5351
194 AACATTTAATATATATATATTTGCAATTTTCTTAAAGGAGATACATTTTGTGGA 253
5352 GACATTTGAATATATCTATTTCTTGCACTAGGATATGATTTCTAATTTTCTTGA 5411
254 GAATTAATTTGCACTTTCAAGAGTGTGCGGTATTTTCACTTAATTAATCTCGTTT 313
5412 GATTCGATTTAGTATTTCCAGAGGAGCATTTTACTTCTCAATTAAGTCTCTT 5471
314 GGTATACCTATTTAAGCTAATTAATTAAGAGTTTAAAGATAGCTTTCTATTTTCAAT 373
5472 AGCTATAGTCTAAACCTATTTATCAAGGTATGGAACGGGCTTATCCCACTTTAA 5531
374 CTATCAAAAAAAGACGATTTCTAATTAACGCGTGAACGATGGGAAATATGCAAGTTT 433
5532 AGGAAGTAAGAAATCTCTACTTACAGCAACTTCTGCTGCAAAAAGGTACTGATAG 5591
434 ATTGAATCATATAACAATCAAAAAAATCTTTGCAATGCTAGTTTATGATACAGA 493
5592 ATTATATGAATCAAAATGAGTTGTGGGAGTTGTAGCCGTCAAGTCTTATGATTAAC 5651
494 AATAGATAAATTAATTAATCATTAACGCTCTATTAATCTGTGGAAGAACCTATAGATT 553
5652 AGATTTTCACATGATTAATTAAGAGT-----AGCAGAGAGAGATAGTAACCTT 5705
554 TTCAACAAGGAGAGTGTGCAACGCTTTTATTAATCTACAAATGAGTTTATAGCT 613
5706 TGGACATCATGAGTGTGATGAGTATTAATCAATCTTCAAGTGAATAAATCAATAT 5765
614 AAAGCAATTCGTTTCAGATTTGAGTTGATGATGATGATGATGATGATGATGATGAT 673
5766 TGGAGAGCTTGTCTCTCAGTTTGAACGATGGAAATGATGATGATGATGATGATGAT 5825
674 ATTGCGTTTACTGCGTTGAAAAACAATAATCCAACTGATGATGATGATGATGATGAT 733
5826 TTTTATATGATTTGGCAGTAAACAAGCAATCCGATGATGATGATGATGATGATGAT 5885
734 AACTTTTTCAGAAATTTTATAGCTATGATGATGATGATGATGATGATGATGATGAT 793
5886 GACTTTTTCAGAAATTTTATAGCTATGATGATGATGATGATGATGATGATGATGAT 5945
794 ACTGAGCGGTAGTGGGTTAATTAATTTGATGATGATGATGATGATGATGATGAT 853

```

```

Db
5946 CGTGGGTCATTTGATGAGGCTGATTAATGATGATGATGATGATGATGATGATGAT 6005
854 TATTCGTAGAGATGATGAGCCGCTATTTTGTCTAGAAAAGATTTGACAGATGAGAC 913
6006 GATTCGAAAGATGAGGCGCTTATCTATTTTGTCTCAGACCGCTATAGAAAAATGCTG 6065
914 CATATTTAATCTTACAGTTTTCATGATGATGATGATGATGATGATGATGATGATGAT 973
6066 TCAGTTCACTTTTATTAAGTTTGTCTATATGATGATGATGATGATGATGATGATGAT 6125
974 CTTCCTCAGCAAAACAGATGAGGCGGTATGATTTTAAATGGAATAACGATCTT 1033
6126 ACTCATGAGAACAAAATACATGAT-----GGGAGATGATTTAAGTGTGCGATGATCT 6180
1034 AGAATTATCTCCATTTGACATTTTATACGCAAAAACAGTTTATGAGATTTACAGATT 1093
6181 CGATACAGAAATTTGTGTTTATACG- GAAAGCTTATGTTGAGAGCTTACACAGTT 6239
1094 TTATATGTTTATTAATGAGGATATGAGTCTAGTTGATACAGTCTACAGTTGATGAT 1153
6240 TTATATGTTTCTAAGGAGATATGATTTGTTGTCACACGCGCTACACAGTGTGACGA 6299
1154 ATTGAAAAATATATCTCTGCTCAAAAGAGACGATGATGATTTAAACAGGATTAACAG 1213
6300 GTATGAACACTATACCCAGAACAAACGTCGCTAAGTTTAACTGTGATTAACAG 6359
1214 TCTGCGCAGGTTAGTGTCTGATATATACAGCTTCGAGAGAGTGTGCTGATGAT 1273
6360 CTATGCGCAGGTAGTGTGAGAGAGATTAAGATTTGATGAGAGTGTCAAAATTAAG 6419
1274 CTATGACATCATTTGATTAATTTGATCTGCTGATGATTTTAAATTTTAAAGACAT 1333
6420 TGTAGCTATATAGAGATTTGACATTTTGAAGAAGCATTTGAATTTTATGAAGACAT 6479
1334 GAAAGTTGATTTGATGAGAGAGGAGATTAAGTAATATGATTAAGTTTGTGCTGCG 1393
6480 TAAAGTTGATTTATAGAGAGCGAGCGAAGT---AGAAATGAATAATATGTTAGTGG 6536
1394 TTTTGAAGGAGCATTTGATCTCATTTTGTATTTTAAACGTTTGTGAAGAGAAAGA 1453
6537 ATCTAGTGAAGGTCATCTCATTTTATATCTTTTAAACCTTTGTGAAAGATTAAGA 6596
1454 AGCTTTTGGGTAACTTTGATTAAGAGATGCAAGCTTTTGAAGATTAAGAAAT 1513
6597 AAGATTTGGGTACTTTGATTAAGCGAGCGCTCGAGTGTACTCTGAGAGAGTGT 6656
1514 GTATCCATGTTACTTCCAACAATGCAATCTCATTTAATTTAGTAAATACTTTCTT 1573
6657 ATATCCCTGTTATATCCAAACATGTAATTAATAAATTAATAAATACTTTCTT 6716
1574 AGCTTCAAAATTTTACGTATGAGAAACCAAGATTTTATTTTCACTGTGCGGCGCT 1633
6717 CGCTATTAATAATTTTAAAAAAGAGACAGATTTGATTTTCTTGTGTCGCTAT 6776
1634 TGTGTCCCTCTTTTACATGAGAAACATTTTGGAGCAAGACATTTATTTTGAAGT 1693
6777 CGCTGTCTCTTTTATATTTGGAAGTATTTGATTAAGAAACAGTGTATTAAGAGT 6836
1694 ATTGATGAGTATTAATTAATCTAATTAATCTGAAAACTAGTTATCCCGTAAACATAT 1753
6837 TTTGATTAAGATTTGATTTCAACCTGACAGGAAAACTTTGTTATCCAGTAAACGATG 6896
1754 TTTTATTTGCTGAGTGAAGAAATGAAGAGGATATCTTAATCTAATCTTTGGGAG 1813
6897 GTTATTTGTCATGAGGAAATGAAGAAATGAAGATTAATCAAGCAATTAATCTCGAG 6956
1814 TATTTTATGATTTTGTGAACATGAGAACTCATGAAACAGTTTATGATGATGAT 1873
6957 GATTTTCTAATGATTTTGTGACAGTATGATCTCATGAAACAGTTTATGATGAT 7016
1874 AAGAGATTTGATTTATGAATAAATGAAGATTAACGAGAAATATTTATTAACA 1933
7017 AAGAGATTTGATTTAAGTGAAGGATTTATTCAGAGATGATTTTATTAACA 7076

```


misc_feature /locus_tag="SPC15A_0004"
/note="3 probable transmembrane helices predicted for
SPC0676 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
1409..1753
/gene="wzg"
/locus_tag="SPC15A_0004"
/note="HMMFam hit to PF02916, DNA polymerase processivity
factor, score 5.4e-63"
1931..2374
/gene="wzg"
/locus_tag="SPC15A_0004"
/note="HMMFam hit to PF03816, Cei1 envelope-related
transcriptional attenuator, score 3.2e-70"
2649..3380
/gene="wzh"
/locus_tag="SPC15A_0005"
2649..3380
/gene="wzh"
/locus_tag="SPC15A_0005"
/EC_number="3.1.3.48"
/note="member of homology group 0001 90"
/codon_start=1
/transl_table=1
/product="protein-tyrosine phosphatase Wzh"
/protein_id="CAI33337.1"
/db_xref="GI:68643022"
/translation="MIDVSHIYFDVDDGKRSREKALLAESYRGQVITVTSNR
KMFETPEEKIAENFLQVRIEAVASDLVIAAGAEIYTPVDVLDKLEKRIPLNDS
RYALIESNMTPRDIHSALSKIMLGITVIAHIEYDALENNEKRVRLIMGCT
QVNSHVLSKLEFGERYKFMKRAQYFLEODLVHVIASDMHNLDRPPHAAEYDLVT
OKYGEAKAOELFDNPRIKIMDOLI"
2961..3275
/gene="wzh"
/locus_tag="SPC15A_0005"
/note="HMMFam hit to PF02811, PHP, C-terminal, score
1.3e-12"
3389..4081
/gene="wzd"
/locus_tag="SPC15A_0006"
3389..4081
/gene="wzd"
/locus_tag="SPC15A_0006"
/note="member of homology group 0002 90"
/codon_start=1
/transl_table=1
/product="capsular polysaccharide biosynthesis protein
Wzd"
/protein_id="CAI33338.1"
/db_xref="GI:68643023"
/translation="MEBONLRIIDVQLPSIMKRLVILVLISSVAFASTVI
KEPFTSTRIYVNRDQEKSLTMDLQGTLYVKDYREITISQVLEBVSDDLKD
LTPKGLANKIKVTPVDTRIYSVXDKQPEASRIANSIREVAEKIVATVSDVT
LTSEARPATPPSPNVRNTLFGFLGAVTVIAVLIELDRVXKPEDVEDVLQIP
LIGLVPDLDMKK"
3407..3823
/gene="wzd"
/locus_tag="SPC15A_0006"
/note="HMMFam hit to PF02706, Lipopolysaccharide
biosynthesis, score 1.6e-52"
join(3455..3523,3923..3991)
/gene="wzd"
/locus_tag="SPC15A_0006"
/note="2 probable transmembrane helices predicted for
SPC0678 by TMHMM2.0 at aa 23-45 and 179-201"
3929..4006
/gene="wzd"
/locus_tag="SPC15A_0006"
/note="TPS00217 Sugar transport proteins signature 2."
4091..4780
/gene="wze"
/locus_tag="SPC15A_0007"
4091..4780
/gene="wze"
/locus_tag="SPC15A_0007"

Query Match 12.4%; Score 866.2; DB 1; Length 18517;
Best Local Similarity 61.9%; Pred. No. 1.2e-78;
Matches 1481; Conservative 0; Mismatches 888; Indels 22; Gaps 6;
CDS /locus_tag="SPC15A_0007"
/EC_number="2.7.1.112"
/note="member of homology group 0003 90"
/codon_start=1
/transl_table=1
/product="tyrosine-protein kinase Wze"
/protein_id="CAI33339.1"
/db_xref="GI:68643024"
/translation="MPTLEISQAKLDPEVKAERYNALCTNLOSGDLKVFSTISYK
LGEKSTSTNTIAMAFARAGYKTLIDGRNSVMGVFAARDKLTGLTEFLSGTDL
SGLCDTNLENLFVIOAGSYSPNPTALOSKNSYMLETRKFDIYIDTPAPGVYI
DAALITRKCDASLTIVAEGRINRDIQAKOELEHTEKPLGVANKEFTSVDRKGSY
GNYGDKKMK"
4795..6162
/gene="wcha"
/locus_tag="SPC15A_0008"
4795..6162
gene /locus_tag="SPC15A_0008"
4795..6162
CDS /locus_tag="SPC15A_0008"
4795..6162
14 ATGGCATTATTTGATATGATACAGTTGCAATTTCTGCAATCTTAACAAGTCATATACC 73
4821 ATTGGCATTATTCAGAGTTTCTGTTATTTATGACTTACTTACTTGTGCTGAG 4880
74 AATGCGATTATTAATCGTCTGGAATTTTATCATATGATGGTTCATATTGGCATT 133
4881 AGAAGCGGAATGTTTCAACAACAGTATGCACTTAATATCTCCATTAATTTGCTT 4940
134 TTTTATATCGATGCGATGCACTTGAATTTGATATGAGTAACTGATAGTTTGAAA 193
4941 TTATATCAGGATTTATGAGACAGATTTCTTAAAGGGAATTTGATGATGCTGCA 5000
194 AACATTAACTATGATATTAATTTGCAATTTTCTTACGCGATATCAATTTTGTGGA 253
5001 GACATTGAATATATATCTTATCTTGGCAGTACGATTAATTTCTAATTTTCTTGA 5060
254 GAATTAATTTGCACTTCAAGCGTGGCGGTATTCATTAATAAATCTCGTTT 313
5061 GGATCGATTATGATTTTCCAGACGAGCATGATTTACTTCCATTAACATGCTCTCT 5120
314 GGTATACCTATTTAAGTATTTTAAAGATTAAGATAGCTTCTATTTTCGCAAT 373
5121 AGCTATGCTTAACTATTTATCAAGTGATTTGGAACGGGCTTATCCCAACTTAA 5180
374 CTATCAAAAAAGACATTTCTAATTAACAAGCGCTGAACGATGGAAAAATATGCAAGTTT 433
5181 AGAAGTAAAGATTTCTTACTTACAGCACTTCCGTCGTAAGAAAGTACTGATAG 5240
434 ATTGGAATCAATAAACAATTTCAAAAAATCTTTGTCATTTGGTATGTTAGTACAGA 493
5241 ATTAATAGATCAAAAGAGTTTGGGAAGTTGTAACCGTCAATGCTTAAATTAACC 5300
494 AATAGATTAATTAATTTATCATTAACGCTCTATTAATTCGTGGAAGACCTATAGGT 553
5301 AGATTTTCAAGATTTGTTTAAAGTAGT-----AGCAGAGGGAGATAGTAACTT 5354
554 TTCAACAAGGAAGTGTGACACAGCTTTTATTAATCTACAAAGTAGTTTTAGACGT 613
5355 TGGAGCATCAGAGTGGTGTGATGAAGCTTTATCAATCTTCCAAAGTAAAAATACATAT 5414
614 AAAGCAATTTGTTCAATTTTGAATTTGATTTGATTTGATTTGATTTGATTTGATTT 673
5415 TGGAGAGCTTGTCTTCAAGTTTAAAGATGGGAATTTGATTTGATTTGATTTGATTT 5474
674 ATTGCTTTTACGCTGTTGAAAAAATTTCAACTCTAGTGAACCATGATTTGATTT 733
5475 TTTTGAATCGTAGTTTGGACATTAACAAGCAAAATCCGAGATGACGAGATTTAAACGTT 5534
734 AACTTTTTCACAAATTTTATTAAGCTTATGATCATATCATGATGAACGACTTTTGATAT 793
5535 GACTTTTTCACAAATTTTATTAAGCTTATGATCATATGATGATGATGATTTGATAT 5594

QY 794 ACTCGAGCGGTAGTGGGTTAATTATTTGGTATAGTTTCTATTTTGTAGTTCCAA 853
 DB 5595 CGTGGGTGATGTGATGGGCTGATATGTGTCTAGCCAGTATGTACTGGTCTTT 5654
 QY 854 TATGTGATAGATGTGACCGGCTATTTTGTCTCAAAAAGATTGACAGATGAGAC 913
 DB 5655 GATTGAAAGATGGGGCTCTACTATTTTGTCTCAGACCGCTATAGAAAAATTTGTGTG 5714
 QY 914 CATATTTAATCTACAAGTTTCGATCGATGTATGTGATGTGAGAGAGCGCAAAAAGA 973
 DB 5715 TCAGTTCACCTTTTATTAAGTTTGCCTATGTGTGATGATGCGAGCGCAAAAAGAAG 5774
 QY 974 CTGTGTCAGCAAAACCAATGCAAGGGGTGATGTTTAAATGGGAAAAACGATCTCT 1033
 DB 5775 ACTCATGGAACAAATACCATGCA-----GGGTGGAATGTTTAAAGGTGACATGATCTCT 5829
 QY 1034 AGAATTACTCCAAATGGAATTCATACGCAAAAACAGTTTAAAGAGTTTACCAAGTT 1093
 DB 5830 CGTATCACGAAATTTGGTGTGTTTATACG-GAAGACTAGCTTGACGAGCTTACACAGTT 5888
 QY 1094 TTATTAATGTTTAAATGGCATATGAGTCTAGTTGTCACGCTCCACCTTACAGTTATGA 1153
 DB 5889 TTATTAATGTTCTAAAGGAGATATGAGTTGTTGGACACGGCCCAACAGGTGACGA 5948
 QY 1154 ATTTGAAAAATATATCTCTGTCTGTCAAAAAGAGAGATTTGATTTTAAACAGGATTTACAG 1213
 DB 5949 GTATGAACTATATCCCGAACAACAAACGTCGGCTAAAGTTTAACTGTATTAACAG 6008
 QY 1214 TCTCTGGCAGTTATGAGTGTGATGATATATCACAGCTTGCAGCAGTGTGCTGTTGA 1273
 DB 6009 CTATGGCAGGTGAGTGAAGAGTGAAGTTAAGAAATTTGATGAAGTTGTCAAAATAGA 6068
 QY 1274 CTATGACATACATTTGATTTGAGTATCTGTGATGATTTAAATTTTATTAAGACAT 1333
 DB 6069 TGTAGCTATATGACATTTGACAAATTTGAAAAGACATTTGATTTATTAAGACAT 6128
 QY 1334 GAAATGTTATTTGTTGAGAGAGAGAAAGTAAATATATGAAAGTTTGTGTGCG 1393
 DB 6129 TAAAGTTTATTTATGAGAGACGAGAGAGT---AGGAATGAAATATGTTTATGTGG 6185
 QY 1394 TTCTTCAAGGGGACATTTGACTCATTTGATTTGTTAAACCGTTTGGAAAGAGAGA 1453
 DB 6186 ATCTAGTGAAGGTGATCTGACTCATTTATATCTTTTAAACCTTTGAAAAGATAAGA 6245
 QY 1454 ACCTTTTGGGTACATTTGATTAAGAGATGCAAGAGCTTTTGAAGATGAAAAAT 1513
 DB 6246 AAGATTTTGGGTACTTTGATTAAGGCGGACGCTCGAGTGTACTCTGTGAGGAATGT 6305
 QY 1514 GTATTCATGTTACTTTCCAACAATTCGAATTCATTAATTTAGTGAATACTTTCTT 1573
 DB 6306 ATATCCCTGTATTTATCCAAACATTCGTAATATTAATAAATTTATTAATAAATCTTTCT 6365
 QY 1574 AGCTTCAAAATTTTACGTGATGAGAAACAGATGTTATTAATTTTCTGTGTGCGCGT 1633
 DB 6366 CGCTATTAATAATTTTAAAAAAGAGACGAGATTGATTTCTTGTGTGCGCGTAT 6425
 QY 1634 TGTGTGCCCTTTCTTTTACATCGAAAACTTTTGGAGCAAAAGCATTTATATTGAGT 1693
 DB 6426 CGCTGTCTTTCTTTTATATTTGGAAATTTTGGATCAAAAGACAGTATATATAGAGT 6485
 QY 1694 ATTGATCGAGTTAATAATCTACATTAACGAAAACTAGTTTATCCCGTACAGATAT 1753
 DB 6486 TTTTGATAGATTAATGATCTTAACACGACGAGAAACCTTTTATTCAGTAACGATAG 6545
 QY 1754 TTTTATTTTCACTGCGAGAAATGAAGAAAGTATATCTTAATCTTAATCTTGGGAG 1813
 DB 6546 GTTATATGTTCAATGGGAAGAAATGAAGAAAGTATACCAAAAGGAATTAATCTCGAGG 6605
 QY 1814 TATTTTATGATTTTGTGAACGTAGAACTCATGAACAAGTTTATTCGATGATA 1873
 DB 6606 GATTTCTCAATGATTTTGTGACAGTAGTACTCATGAACAACAGTTTAAATGCACTATT 6665

QY 1874 AAGAGATTGATTATTTGAAAAAATGGAAGTATATACCGAGCAATATTATTCAAACA 1933
 DB 6666 AAAAGGTTGATGATTTAAAGGTAGGATTTTATTCAGATGAAGTTTATTCAAACA 6725
 QY 1934 GGATATTTGATATTTTCCAGATATTTGCAAGTATTAATAAATTTCTCAGTTACAAAGA 1993
 DB 6726 GGAATCTTAATTAATGTTCCAAATTTTGTAAATGGGAAATTTATATCTTATGAAAA 6785
 QY 1994 ATGGAACAATATATTAACAATCAGAGTATTTTGTGACAGGAGGCCCGCTACTTT 2053
 DB 6786 ATGAATCAATTTATGAGGAGCTGATACATCATTAACCATGCGGTGACCTACGTT 6845
 QY 2054 ATGAATTCATTTCCAAAGAAAAAACAATTTGTTTCTTACGCAAAAAGTATGCT 2113
 DB 6846 ATGGCAGTATTTGCTTAAGGAAAAAGTCCGATATTTGTTCCAGATTTAAAAAATTTGCT 6905
 QY 2114 GAACATGTAATGATCATCAGTATGATTTGTAAGAAGATTTTCAAGATATATAT 2173
 DB 6906 GAGCATGTAGACCACTCATCTGAAATTTGTTGAGAAAGTGTAAATGCTATATTTG 6965
 QY 2174 TATTTATGAAAAATATAGATGATTTGTTGAAAAAATTTATGAAGTTCTTACG----- 2228
 DB 6966 ACAGTATCATTAACATTAAGTATCTTAATTCCTATATTTCTAATTTGATGAGCGAAAA 7025
 QY 2229 -AACTACTTTTACATCAATATATATTTTGTGAAAGATTAACAATATGTTGA 2287
 DB 7026 AATTCAGATTTGAAAGCAAAATATCTATTTATGAGAAATTTATGATGATTAAT 7085
 QY 2288 AATTTAATGAGATCAAGAAAAATGAATATTAATAAAGATGCTAT-TTGATTAATGCTT 2346
 DB 7086 CAGTATGAGAGCAGGTGATCATGATGATGATGATGATGATGATGATGATGATGAT 7145
 QY 2347 ATCATATTTTCTCAGATTTTACTGAGAGGATTAACATATTAATCATCT 2397
 DB 7146 GTTATTCATATATTCGTTATTTAGTGAATTTGCTATCTATATTAATTTCTGT 7196

RESULT 12
 CR931661
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Streptococcus pneumoniae
 Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 17717)
 CR931661 17717 bp DNA linear BCT 30-JUN-2005
 Streptococcus pneumoniae strain 34357 (serotype 13).
 CR931661 GI:68642970

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL
 COMMENT

1 (bases 1 to 17717)
 Bentley, S.D., Aarensen, D., Mavroidi, A., Saunders, D.,
 Rabinowitch, B., Collins, M., Donaghy, K., Harris, D., Katoft, M.S.,
 Murphy, L., Quail, M.A., Samuel, G., Skovsted, I.C., Barrell, B.G.,
 Reeves, P., Parkhill, J. and Spratt, B.G.
 Genetic analysis of the capsular biosynthetic locus from all 90
 serotypes of Streptococcus pneumoniae
 Unpublished
 2 (bases 1 to 17717)
 Bentley, S.D.
 Direct Submission
 Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
 Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
 NOTE: This sequence was generated from a PCR product representing
 the region from dexB to alyA and is not necessarily responsible for
 the expressed capsule serotype. For a detailed description of how
 CDS products were predicted see the associated publication.
 Location/Qualifiers
 1..17717
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /strain="34357"
 /db_xref="taxon:1313"
 <1..1313
 gene

```

/gene="dexB"
/locus tag="SPC13_0001"
<1. 134
/gene="dexB"
/locus tag="SPC13_0001"
/codon start=3
/translation="CA133294.1"
/db_xref="GI:68642971"
/translation="SNBBDLYEGKVKSLIENTLAQEVFEKQILVPMDFVELL"
293. 466
/gene="a11B"
/locus tag="SPC13_0002"
/pseudo
293. 466
/gene="a11B"
/locus tag="SPC13_0002"
/pseudo
/codon start=1
/translation="putative oligopeptide-binding protein A11B
(pseudogene)"
/db_xref="PSEUDO:CA133295.1"
complement(join(459. 995,997. 1323))
/gene="tnp"
/locus tag="SPC13_0003"
/pseudo
complement(join(459. 995,997. 1323))
/gene="tnp"
/locus tag="SPC13_0003"
/locus tag="SPC13_0003"
/note="member of homology group 0004 74"
/pseudo
/codon start=1
/translation="putative IS630-Spn1 transposase"
complement(934. 1305)
/gene="tnp"
/locus tag="SPC13_0003"
/note="HMPfam hit to PF01710, Transposase, Synchocystis
PCC 6803, score 3.9e-23"
/pseudo
1536. 2981
/gene="wzg"
/locus tag="SPC13_0004"
1536. 2981
/gene="wzg"
/locus tag="SPC13_0004"
/note="member of homology group 0000 90"
/codon start=1
/translation="integral membrane regulatory protein Wzg"
/protein_id="CA133297.1"
/db_xref="GI:68642972"
/translation="MSRRFKKSGSOKKRSVNIITLYLLVCFLFLFKYNILAF
RYNIIVTALVILVALVGLLITIKKAEKRTIFLLFSILSVSLFAVOQFGLNTR
LNATSVYSYSISVVVLADSDIENVQLSVTPAPCTNDNENIKLADIKSSNPLT
VNQSSYLAIVKSLIAGETKALVANSVFENITELRYPNASKIKTKIYTGFTKVPAP
KTSKQSFNIYVSGIDTYGPISSVSRSDVNIITVARDTKILLITTPDAVVPADG
GNQKDKLTAGIYGVDSIHTELENLYGVNIVYVRLNFTSLKMTIDLGGVDVNDG
EBSALHGKFFYPGVGNVHLDSEQALGVRERSLADGRDRGRQCKIVAILQKLTST
EALKVSTIINSLODSIQTNMPLTMINIVNQLSESGNKKVNSDCLKGGRDLPSY
AMPDSNLYMEIDSSLAIVKAIQDMEGR"
1536. 1652
/gene="wzg"
/locus tag="SPC13_0004"
/note="Signal peptide predicted for SPC0636 by SignalP 2.0
HMM (Signal peptide probability 0.993) with cleavage site
probability 0.538 between residues 39 and 40"
join(1584. 1643,1671. 1739,1758. 1826)
/gene="wzg"
/locus tag="SPC13_0004"
/note="3 probable transmembrane helices predicted for
SPC0636 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"

misc_feature
1743. 2087
/gene="wzg"
/locus tag="SPC13_0004"
/note="HMPfam hit to PF02916, DNA polymerase processivity
factor, score 1.1e-63"
2265. 2708
/gene="wzg"
/locus tag="SPC13_0004"
/note="HMPfam hit to PF03816, Cell envelope-related
transcriptional attenuator, score 1.4e-69"
2983. 3714
/gene="wzh"
/locus tag="SPC13_0005"
2983. 3714
/gene="wzh"
/locus tag="SPC13_0005"
/BC_number="3.1.3.48"
/note="member of homology group 0001 90"
/codon start=1
/translation="protein-tyrosine phosphatase Wzh"
/protein_id="CA133298.1"
/db_xref="GI:68642973"
/translation="MIDISHIVFDVDDGPKSIESKALLREAYNQGVMIVSTSHR
KGMFEPBEKIATNFIKREIADKADVIAGAIYITLDALEKLEKEIPTLNS
RYALIEFSMHTSYREIHAGLSNIMGITPVIAHIERYDALENNERVREILIMGCT
QINSYHVLKDPKPGCEKCKEMKGRARYFLERDLVHVVASDMHNIDSRPPNQAYDII
KRYGAKKAEFFVDNPKIIMDLT"
3295. 3609
/gene="wzh"
/locus tag="SPC13_0005"
/note="HMPfam hit to PF02811, PHP, C-terminal, score
2.3e-10"
3723. 4415
/gene="wzd"
/locus tag="SPC13_0006"
3723. 4415
/gene="wzd"
/locus tag="SPC13_0006"
/note="member of homology group 0002 90"
/codon start=1
/translation="capsular polysaccharide biosynthesis protein
Wzd"
/protein_id="CA133299.1"
/db_xref="GI:68642974"
/translation="MKQNTLEIDVQLFRTLMKRRKVLILVAITSSVAFAVSTFYI
KBEFTSTRIYVNNRNOEKSGITNODLAGSYLVQYREIILSDVLEVISDLKID
LTPKGLANKIKTVPDTRIVSVNDRVBEASRLANSREVAQKIIISITVSDVT
TLSEARPAISPSSPNKKRNTLIGFLAGIGTSIVTFLLELDTHVRPEDIEDTLQWT
LIGVNPNSKLR"
3741. 4157
/gene="wzd"
/locus tag="SPC13_0006"
/note="HMPfam hit to PF02706, Lipopolysaccharide
biosynthesis, score 6.8e-54"
3789. 3854
/gene="wzd"
/locus tag="SPC13_0006"
/note="1 probable transmembrane helix predicted for
SPC0638 by TMHMM2.0 at aa 23-44"
4263. 4340
/gene="wzd"
/locus tag="SPC13_0006"
/note="PS00217 Sugar transport proteins signature 2."
4425. 5108
/gene="wze"
/locus tag="SPC13_0007"
4425. 5108
/gene="wze"
/locus tag="SPC13_0007"
/BC_number="2.7.1.112"
/note="member of homology group 0003 90"

```

```
/codon_start=1
/translation_table=11
/product="tyrosine-protein kinase wze"
/protein_id="CAI3300.1"
/db_xref="GI:68642975"
/translation="MPTLEIAQKLEFIKKEEYNNALCTNIQLSGDKLVISVSVN
PDEGKTTTSVNSPARGAKYKTLIDGTRNSVISGKSEKFTGLTEPLSGADL
SHGLCDTNIENLFIIVOSGSVSPNALQGNFMDIETLRKFPYIIIVDTPIGIVI
DAIITTKDCDASILVTATSEVNVKRVQAKQOLEQTKLFLGIILINKPVDHKTKGSY
GDYGVYKXK"
5124..6491
/locus_tag="SPC13_0008"
/gene="wcha"
/locus="wcha"
/locus_tag="SPC13_0008"
/notes="member of homology group 0005 66"
/codon_start=1
/translation_table=11
/product="undecaprenylphosphate glucosylphosphotransferase"

Query Match      12.4%; Score 865.2; DB 1; Length 17717;
Best Local Similarity 62.3%; Pred. No. 1.6e-78;
Matches 1432; Conservative 0; Mismatches 853; Indels 15; Gaps 4;

QY      16  TGGCATTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAACAGATCATATACCA 75
DB      5152 TGGCAATATTTACAGAGATATATTTGCAATTTATTTGTTATTTCTTACGTTGTTAGAG 5211
QY      76  ATGCTGATTTAAATCGTCTGCAATTTTATCATTAATGATGCTCATTTATTTTGCATTTT 135
DB      5212 AGACAGACTTGAACGTTCTTGATATGATTAATACCTTCCACTTTTGTGATTTCT 5271
QY      136  TTATATCTGATGCCAGTTGAATTTGAGTATAGAGTAACTCGATAGAGTTTGAAGAAA 195
DB      5272 ATTTTAGTTCCTATGTAATATTTTAAAAAGAGGCACTTAGTGAATTTAAATAGTA 5331
QY      196  CATTTAACATATATATATTTTGCATTTTCTTACGAGATATCATTTTGTGGAGA 255
DB      5332 CTATAGATATATATTTTCTTGGCAATGCTAATAGTATTAATCTTTTATATACGG 5391
QY      256  ATATATTCGCACTTCAAGACGTGTCGCTGATATTTCACTTAATTAACCTTGTTTGG 315
DB      5392 AACGTTTATGATCTCTAGAGAGAGATGATATCTTCTTAACCTTGAAGAAATCCT 5451
QY      316  TATACCTATTTAAAGTATTTAAGCAGTTTAAAGATGCTTTCTATTTTCCAAATCT 375
DB      5452 TATACCTGTTAAATTTCTAGTAAAGAAATATGGAAGCATGTGTTTTTAAATCTAAAA 5511
QY      376  ATCAAAAAAAGACATCTTAATTAACAACGCTGAACAGATGGGAAAAATATGCAATTTAT 435
DB      5512 ATAGCAGAAAAATTTTATCTGTTAACAGTAAAGAAAAATATGAAAAATCTTGATTAAT 5571
QY      436  TTGAATCAGATTAACAATTAACAAAAATCTTGTGCAATGATGTTATTTAGTACAGAAA 495
DB      5572 TGCTAGATCTGATGAACCTTTCATGAAATTTGGTAGAGTAAAGTTTGGATTAATCTG 5631
QY      496  TAGATTAATTAATTTATCATTTACCGCTATATTTCTGTGGAAGAAAGCTATAGAGTTT 555
DB      5632 ATTTTCAACATGATTAATATCTGTAATGAAAA-----GGAATAATTAATGAATTTG 5685
QY      556  CAACAGAGAAAGTGTGACACAGCTCTTATTAATCTTACCAAGTGAAGTTTATACGTAA 615
DB      5686 CAACGCAATGAAGTGTGATGAGGTGTTTGCATCTTCCAGGGAGAGGCTACATATG 5745
QY      616  AGCAATTCGTTTCAGATTTGAGTTGATGATTTAGTATTTAGAGGCTGATATTAATCAT 675
DB      5746 GAGAAATTAATCTTAGGTTTGAACATGGGATAGATGTAATCTTAATCTTAAAGCAT 5805
QY      676  TCGGTTTACTGCGTTGAAAAAACAATAATCAACGCTAGAGTGAACCATAGCATTTGTA 735
DB      5806 TTGATTAAGAAATTTGGTGCATTAACAATTAATCATGATGATGATGATGATGATGATG 5865
```

```
QY      736  CTTTTCACAAATTTTATTAAGCTAGTCAATATCATATGAAACGACTTTTGATATAC 795
DB      5866 CTTTCTCTCAAAATTTTATTAATACTAGTCAATGATTTCAAGAAATTTTCGATATTT 5925
QY      796  TCGAGCGGATGTCGGGTTAATTAATTTGNGATATAGTTTCTATTTTGTAGTTCCAAATTA 855
DB      5926 GTGGGCACTATTTGCTTTGTTTGTCTATAGCTAGTCAATTTTATGTTCCATTTGA 5985
QY      856  TTGATGAGATGATGAGACCGGCTATTTTGTCTAGAAAAAGAGTTGAGCAAGATGACGCA 915
DB      5986 TTGATGAGATGAGCGGACGAGCTATTTTGTCTCAAACTCGATATGAGAAAAATGCTGCAC 6045
QY      916  TATTTACATTTCAACAAGTTTGCATGATGATATGTTGATGCTGAGAGAGCGCAAAAAAGACT 975
DB      6046 ATTTTACCTTTTATTAATTTCCCTTTCATGAGCGATGATGCTGAACCTATCAAGAAACAT 6105
QY      976  TGCTGAGCAAAACAGATGCAAGGATGAGTATGTTTAAATGAGAAAAACGATCTTAG 1035
DB      6106 TATGAGATCAAAATTCGATGCAAGGATGATGTT-----AGATGCAATATGATCTCG 6160
QY      1036  AATTACTCCAAATTTGACATTTTCATACGCAAAAAACAATTTAGACAGATTTACACAGTTT 1095
DB      6161 TGTTCAAAAAATTTGATCGCTTATTCG--TAAACAGATTGGATGATATGCGCAGTTT 6219
QY      1096  ATATATGTTTAAATTTGCGATATGATCTATGTTGTAACGCTCACCTACAGTTGATGAT 1155
DB      6220 GGAATGCTTTTATGAGATATGATGTTGTTGTTGAGAACAGTCCACTCAAGTAGACAGT 6279
QY      1156  TTGAAAAATATATCTCTGCTGCAAAAGACAGATGATGATTTTAAACAGAGATTTACAGTTC 1215
DB      6280 ATGATCAGATATCTCAGAACAGAAACGTCGACTCAGCTTTAAACCTGATATACAGTT 6339
QY      1216  TCTGCAAGGTTAGTGTCTGATTAATATCAACAAGCTTTCAGACAGTATGTTGCTGACT 1275
DB      6340 TATGCAAGTTACCGGCGCTATGTAATAACGATTTTACAGTGTGTTAAATATGATG 6399
QY      1276  TAGCATATATGATTAATTTGACTATCTGCTCAGATTTTAAATTTTATTAAGACAGTGA 1335
DB      6400 TGCTTATATGATTAATTTGACATCTGAAAAAGATTTTAAATTTTGTCTTAAACGTTGA 6459
QY      1336  AAGTTGATTTGTTGAGAGAGAGAAAGTAAAGTAAATGAAAGTTTGTGCTGCTT 1395
DB      6460 AAGTTGATTTTAAAGAGATGAGCGAAGTAA--GGAATGAAAGTATGCTAGTTGAT 6516
QY      1396  CTTCAAGGAGCAATTTGACTCACTTGTATTTTAAACCGTTTGAAGAAAGAAAGAAC 1455
DB      6517 CTAGTGAAGCCATTTTATCTCATTTGATTTATTAACCTTTCTGAAAGATTAAGAAA 6576
QY      1456  GTTTTGGGTATGATTAAGAGAGATGCAAGAACTTTTGAAGAAAGAAATATGTC 1515
DB      6577 GATTTGGGTATCTTTGATTAAGAAAGATGCAAGAAATTTTGGGAAATGAATATTTT 6636
QY      1516  ATCATGTTACTTTCCAAACAAATCGCAATCTCTTAATTTAGGAAAAATATCTTCTTAG 1575
DB      6637 ACTCATGCTATATCTTACTAATCTGATCTTAAAAATCTTATTAATAATATCAATTTCTG 6696
QY      1576  CTTTCAAAATTTTACGATGAGAAACAGATGTTATTTTCAATCTGTCGCGCGCTTG 1635
DB      6697 CTTTCAATATTTTGAAGAAAGAACACCTGATATTTATCGATCATACGAGGAGCTGTAG 6756
QY      1636  CTGTCCCTCTTTTACATGCGAAAACTATTTGAGCAAGACAGATTTATATGAAATAT 1695
DB      6757 CAGTTCTTTCTTTTATCTAGGAAAAATATTTGTTGCTAAGACAGTCTATATGAAATAT 6816
QY      1696  TTGATGAGTTAATTAATCTACATTAATCTGAAAACTAGTTTATCCGTTAAAGATATTT 1755
DB      6817 TTGATGAAATTTGATGCTCCCAAGATGACGGGAAACTTGTATTCAGTAACTGACAGGT 6876
QY      1756  TTATTTGTCAGTGGAGAAATGAAGAGTATATCTTAATCTATTAATCTATGAGGAGTA 1815
DB      6877 TTATTTGTCATGGGAGAAATGAAGAAATGATATACCAAAAGCAATTTTACCTAGGAGGA 6936
QY      1816  TTTTATATGATTTTGTAAACATGAGAACTCATGAACAAAGTTTAAATGATGATTA 1875
```

Df		6937	TATTCTAATGATTTTGTGACGTGGTAGCACATAACAACATTTGATCGATTGATTAA	6996
OY		1876	AGAGATTGATTTTATGAAGAAAAATGAGATTAACCAAGAAATTTATTCAAACAGG	1935
Df		6997	AGAACTGATCGTTTAAAAAAAAGAATCTTATTCAGATGAGTATTTATTCAAACAGG	7056
OY		1936	ATATTCTGACTATATCCAGAAATATTCAGATTAATAAAATTTCTCAGTTACAAAGAAAT	1995
Df		7057	ATATTCCAAAATATATTCGAAATATTTGAGAGGGGAAAAAGATTAATTTCTTATGAAAAAT	7116
OY		1996	GGAACATATATTTAACATACAGAGTAGTATTGTCACGAGGCCCGCTACTTTAT	2055
Df		7117	GAATCAATGATGAGAAATGAGATATTTGCTATTACTCATGCTGACCAAGCAGCGTTAT	7176
OY		2056	GAATTCATTTATCCAAAGAAAAAAACAATTAATTTGTTCTTGACAAAAAAAGTATGCTGA	2115
Df		7177	GGGGGTATTTGCTAAAGAAAAATTCGATAGTTGTTCTTGACAAAAAAATTTGAAGA	7236
OY		2116	ACAATTAATGATCATCAAGTAGTGTGTAAGAAATTTTACAGATTAATATATTTT	2175
Df		7237	GCATGCTGATGACCACTGCTTGAAGATTTCTGATTAAGATTAAGCAGACTTCCCATATAT	7296
OY		2176	AATTATTAAGAAATATAGATGATTTGTTGAAAAAATTTATGAGTTTCTAAGCAAACATA	2235
Df		7297	TGTAATTAATGATTAATAAAATCTAGTATTAATTAAGAAATTAATCAAGAAAC	7356
OY		2236	CTTTCATCATTAATTAATTTTTTTTGTGAAAGATTAACAAATAGTTGAAAAATTTAA	2295
Df		7357	TATAATATGATATATTAAGAGCTTTGTAAGAAAAATTGAGAAAGAAATAGAGTTTCTA	7416
OY		2296	TGAGATCAAGAAATGAAT	2315
Df		7417	ATGAATTAAGAAATCATT	7436
RESULT 13				
LOCUS		CR931664	18624 bp	DNA linear BCT 30-JUN-2005
DEFINITION		CR931664	Streptococcus pneumoniae strain 7904/39 (serotype 15b).	
VERSION		CR931664.1	GI:68643045	
KEYWORDS				
SOURCE				
ORGANISM				Streptococcus pneumoniae Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE				1 (bases 1 to 18624) Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D., Rabinovitch,E., Collins,M., Donaghy,K., Harris,D., Kallioft,M.S., Murphy,L., Quail,M.A., Samuel,G., Skovhede,I.C., Barrell,B.G., Reeves,P., Parkhill,J. and Spratt,B.G. Genetic analysis of the capsular biosynthetic locus from all 90 serotypes of Streptococcus pneumoniae Unpublished 2 (bases 1 to 18624) Bentley,S.D. Direct Submission Submitted (09-DEC-2004) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, E-mail: sdb@sanger.ac.uk NOTE: This sequence was generated from a PCR product representing the region from dexA to alvA and is not necessarily responsible for the expressed capsule serotype. For a detailed description of how CDS products were predicted see the associated publication. Location/Qualifiers 1..18624 /organism="Streptococcus pneumoniae" /mol_type="genomic DNA" /strain="7904/39" /db_xref="taxon:1313" <1..134 /gene="dexB"
FEATURES				
source				
gene				

gene /locus_tag="SPC15B_0004"
/note="HMMpfam hit to PF03816, Cell envelope-related transcriptional attenuator, score 5.2e-70"
2651..3382
/gene="wzh"
/locus_tag="SPC15B_0005"
2651..3382
/gene="wzh"
/locus_tag="SPC15B_0005"
/EC_number="3.1.3.48"
/note="member of homology group 0001 90"
/codon_start=1
/transl_table=11
/product="protein-tyrosine phosphatase wzh"
/protein_id="CAI33360.1"
/db_xref="GI:68643048"
/translation="MIDHSHIVFDVDDPKRSRESKALLAESTYQGVRIIVSTSHR
KMFETPEEKIENFLQVREIAKVASDVIYAGAEIYTPVDLKEKRLPTLND
RYALIEFSWNTPYRDIHSAISKILMIGTIPVIAIERYDALENNERVELIDMGCT
OVNSSHVLPKPLFGERYKEMKRAQYFLERDLVHVASDMNLDSPYMQAYDIIA
KRYGAKKAEPEVDNPKTIMDQI"
2963..3277
/gene="wzh"
/locus_tag="SPC15B_0005"
/note="HMMpfam hit to PF02811, PHP, C-terminal, score
3.5e-13"
3391..4083
/gene="wzd"
/locus_tag="SPC15B_0006"
3391..4083
/gene="wzd"
/locus_tag="SPC15B_0006"
/note="member of homology group 0002 90"
/codon_start=1
/transl_table=11
/product="capsular polysaccharide biosynthesis protein
wzd"
/protein_id="CAI33361.1"
/db_xref="GI:68643049"
/translation="MKEHNTLEIDILQFRALMKKLVILVAITSSVAVSTTVI
KPEFTSTRIIVYVNRNQEKSGLTNDLQAGTVLVKDYBILISQVLEEVSDKLD
LTPKGLANKIKVTPVDTRIIVSVNDRVPEEASRIANSIREVAQKISITRVSQVT
TEBARPAISPSPIKRNILIGLAGVGTVIVLLEHLDHVRPRDIEDTLQMT
LGVVPLSKLK"
3409..3825
/gene="wzd"
/locus_tag="SPC15B_0006"
/note="HMMpfam hit to PF02706, Lipopolysaccharide
biosynthesis, score 8.8e-54"
join(3457..3522,3925..3993)
/gene="wzd"
/locus_tag="SPC15B_0006"
/note="2 probable transmembrane helices predicted for
SPC0700 by TMHMM2.0 at aa 23-44 and 179-201"
3931..4008
/gene="wzd"
/locus_tag="SPC15B_0006"
/note="PS00217 Sugar transport proteins signature 2."
4093..4776
/gene="wze"
/locus_tag="SPC15B_0007"
4093..4776
/gene="wze"
/locus_tag="SPC15B_0007"
/EC_number="2.7.1.112"
/note="member of homology group 0003 90"
/codon_start=1
/transl_table=11
/product="tyrosine-protein kinase wze"
/protein_id="CAI33362.1"
/db_xref="GI:68643050"
/translation="MPTLEIAQKLEFIKKAEEYNAICTNIQSGDKLVISVSVN
PBGSKTTTSVINARSFARAGVKTLLIDGDRNSVMSGFPSREKRTSLTFELSGTADL

SHGLCTNINENLVIVGSGSVSPNFTALQSNPNMDIETLRKVEDYIIVDPAPIGIVI
DAALITKCDASILVITATGEVNRKDVOAKQDLQTKLPLGVILNKPFDVQHKRYGSGY
GDYNNYGRKK"
4792..6159
/gene="wcha"
/locus_tag="SPC15B_0008"
4792..6159
/gene="wcha"
/locus_tag="SPC15B_0008"
/note="member of homology group 0005 66"
/codon_start=1
/transl_table=11
/product="undecaprenylphosphate glucosephosphotransferase
wcha (initial sugar transferase)"
/protein_id="CAI33363.1"
/db_xref="GI:68643051"
/translation="MDEKGLKPLAVLOSIIIVLVVPLSFVRETELSSMVLVYL
HFPVFPYSYGNRPFKRHLYVFNSTRIYFEPFALISVLPFIAERPSIRGAVYF
LTLEGISLYLNPFLVKTKWKFVFRILKNSKILLITVTKMNEKVLDKLJESDEL
SWKL

Query Match 12.2%; Score 851.8; DB 1; Length 18624;
Best Local Similarity 62.2%; Pred. No. 3.4e-77;
Matches 1447; Conservative 0; Mismatches 857; Indels 21; Gaps 6;

QY 16 TGGCATTTATGATATGATAGCATGTCGAATTTGCGCATCTTAACAAGTCATATACCA 75
DB 4820 TGGCAGTATTACAGAGTATTATGTCATTTATGTTATTTCTTGACTTTGTTAGG 4879
QY 76 ATGCTGATTTAAAGCTTCGTGGAATTTTATCATATGATGATGTCATTTTGGCATTTT 135
DB 4880 AAGCAGAACTTGAACGTTCTCGATGTTATATCTTACCTTCCACATTTTGTGATTC 4939
QY 136 TATATCTCGTATGCGAGTGAATTTGAGTATAGGTAATCTGATAGATTGAAAAA 195
DB 4940 ATTTTAGTTCCTATGATATATATTTTAAAGAGGCACTTAGTGATTATAGTA 4999
QY 196 CATTTACTATGATATATATTTGCAATTTTCTACGCGATTCATTTTGTGGAGA 255
DB 5000 CTATTAAGATATATTTTCTTGGCAATGATATAGTATATTAACCTTTTATAGCGG 5059
QY 256 ATTAATTTGCACTTTCAAGACGTGGCGCGGTATTTGCAATTAATTAACCTTCGTTTGG 315
DB 5060 AACGGTTAGTATCTTAAAGAGAGATGATATCTTTTACTTTTAAAGAGAAATTCCT 5119
QY 316 TATACCTATTTAAGTATATTTAAGCAGTTTAAGATAGCTTCTATTTTGCACATCT 375
DB 5120 TATACTGTTAAATTTCTTAGTAAGAAATATTTGAAGACATGTTTTTAATCTTAAAA 5179
QY 376 ATCAAAAAAAGCAGATTTCAATTTAACAAGCGCTGAAGAGAAATATGCAAGTTTAT 435
DB 5180 ATGCAAGAAAAATTTTACTGTTAACGATTAACGAAAAATATGAAAAAGTTCTTGATTAAT 5239
QY 436 TTGAATCATATAACAAATTTCAAAAAAATCTTGTGCAATGATGTTTAAAGACGAAA 495
DB 5240 TGTCAAAATCTGATGAACCTTTCATGGAATTTGTACAGATTAAGTTTGAATTAATCTG 5299
QY 496 TGATTAATTAATTTATCATTAACGCTATATATCTGTGAAGAGCATTAAGCTTT 555
DB 5300 ATTTTCAACATGATTAATAATCTGTAATTTGAAAA-----GGAAAAATATTTGAATTTG 5353
QY 556 CAACAAGGAAGTGTGACCAAGCTTTTATTAATCTCAACAGTGAATTTTGAAGCTAA 615
DB 5354 CAACGATGAAGTTGTGATGAGAGTGTGTGCAATCTTCAGAGAGAGACTACGATATTG 5413
QY 616 AGCAATTCGTTTCAATTTTGAAGTTGTATGATTAATGAAGCGTTGATATTAATTCAT 675
DB 5414 GAGAAATTAATCTGATTTGAACAAATGGGATATGTAATGTAATCTTAAAGCAT 5473
QY 676 TCGGTTTTCTGCGTTGAAAAACAATAATCAACGCTAGGTGACATAGCATTTGTA 735
DB 5474 TTATATGAATTTGGGTGCGAATTAACAATTCATAGATGATGAATGAATGATGTA 5533
QY 736 CTTTTCACAAATTTTATTAAGCTTATGATATCATGATGAAGACATTTTGGATATAC 795

Db 5534 CTTTCCTCAAAATTTTAAAACTAGTCATGATTTCAAAAGAAATTCGAAATTT 5593
 Qy 796 TCGAGAGGAGTGGCGGTTAATTAATTGGGTATAGTTCTATTTTGTAGTCCAAATTA 855
 Db 5594 GTGGTCCACTATTTGGCTTTGTTTGTCTATTAAGTGTAGTTTATGTTCCATTTGA 5653
 Qy 856 TTGGTGAAGATGGTGGACCGGCTATTTTGTCTCAGAAACGAGTTGACAGATGAGCGCA 915
 Db 5654 TTGGTGAAGATGGCGGACCGGCTATTTTGTCTCAGAAACGAGTTGAGGAAAAATGCTGAC 5713
 Qy 916 TATTTACATTTCTACAAAGTTTCAGTCAGATGATGATGATGATGATGATGATGATGAT 975
 Db 5714 ATTTTACCTTTTATTAATTCCTGTCGATGATGATGATGATGATGATGATGATGATGAT 5773
 Qy 976 TGTCTGACCCAAAACGAGTGCAGAGGCGGATGTTTAAATGAGAAAAACGATCTGAG 1035
 Db 5774 TGAATGATCAAAATTAAGATGCAAGGTGTATGTTT-----AGATGACATATATCTCTG 5828
 Qy 1036 AATTAATCAATTTGACATTTTCAAGCAAAACAGTTTACAGATTTACAGATTTT 1095
 Db 5829 TGTTCACAAAATTTGTCCTTTATTCG-TAAACCAAGTTTGAATGATGATGATGATGATGAT 5887
 Qy 1096 ATTAATGTTTTAATTTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1155
 Db 5888 GGAATGCTTTTATAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5947
 Qy 1156 TTGAAAATATATCTCTGCTCAAAAGACGATGATGATGATGATGATGATGATGATGATGAT 1215
 Db 5948 ATGATCAGATATCTCCAGAAACAGAAACGTCAGCTTAAACCTGATTTTACAGATTT 6007
 Qy 1216 TCTGCGAGGTTAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1275
 Db 6008 TATGCGAGGTTAGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6067
 Qy 1276 TAGCATACATTTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335
 Db 6068 TGGCTTATATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6127
 Qy 1336 AAGTTGATTTGTTGAGAGAGGAGTGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1395
 Db 6128 AAGTTGATTTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6184
 Qy 1396 CTTTACGGGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1455
 Db 6185 CTAGTGAAGGCGCTTAACTCATTTGATTTTAAACCTTTCTGGAATGATTAAGAA 6244
 Qy 1456 GTTTTGGGTTAATTTGATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1515
 Db 6245 GATTTTGGGTTAATTTGATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 6304
 Qy 1516 ATTCATGTTAATTTTCAACAAATCGCATCTCAATTAATTAATTAATTAATTAATTAATTAAT 1575
 Db 6305 ATCCCTGTTATTAATTCACCAATCGTATTAATTAATTAATTAATTAATTAATTAATTAAT 6364
 Qy 1576 CTTTCAAAATTTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1635
 Db 6365 CCAATTAATTTTAAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6424
 Qy 1636 CTTTCCCTCTTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1695
 Db 6425 CTTTCCCTCTTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6484
 Qy 1696 TTGATCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1755
 Db 6485 TTGATCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6544
 Qy 1756 TTAATGTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1815
 Db 6545 TTAATGTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6604
 Qy 1816 TTTTAAATGATTTTGTAACTAGTGAATCAATGATGATGATGATGATGATGATGATGATGAT 1875

Db 6605 TATTTAATGATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6664
 Qy 1876 AGAGATGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1935
 Db 6665 AGAGATGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6724
 Qy 1936 ATATTCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1995
 Db 6725 ATATTCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6784
 Qy 1996 GGAACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2055
 Db 6785 GAATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6844
 Qy 2056 GAATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2115
 Db 6845 GGGGGTTATTTGCTTAAAGAAAGAGTTCCGATGATGATGATGATGATGATGATGATGATGAT 6904
 Qy 2116 ACATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2175
 Db 6905 GAT 6964
 Qy 2176 ATTTATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2233
 Db 6965 AGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7024
 Qy 2234 --ACCTTACATCAATTAATTAATTTTGTGAAAGATTTAAAC--AAATGTTGAAAA 2289
 Db 7025 TGGATTTGTTTGAACAAAGAAAGATTTATCTACATTTTAAATGCGGAAATTAATTAAT 7084
 Qy 2290 ATTTATGAGATCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2334
 Db 7085 ATTTGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7129

RESULT 14
 CR931665 18626 bp DNA linear BCT 30-JUN-2005
 LOCUS Streptococcus pneumoniae strain 553/62 (serotype 15c).
 DEFINITION CR931665
 ACCESSION CR931665.1 GI:68643072
 VERSION
 KEYWORDS
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 18626)
 Bentley, S.D., Aanesen, D., Mavroidi, A., Saunders, D.,
 Rabinowitsch, E., Collins, M., Donoghue, K., Harris, D., Kalof, M.S.,
 Murphy, L., Quail, M.A., Samuel, G., Skovsted, I.C., Barrell, B.G.,
 Reeves, P., Parkhill, J., and Spratt, B.G.
 Genetic analysis of the capsular biosynthetic locus from all 90
 serotypes of Streptococcus pneumoniae
 Unpublished
 2 (bases 1 to 18626)
 Bentley, S.D.
 Direct Submission
 Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
 Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, E-mail: sds@sanger.ac.uk
 NOTE: This sequence was generated from a PCR product representing
 the region from dexB to a1a and is not necessarily responsible for
 the expressed capsule serotype. For a detailed description of how
 CDS products were predicted see the associated publication.
 Location/Qualifiers
 1..18626
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /strain="553/62"
 /db_xref="taxon:1313"
 <1..134
 /gene="dexB"
 /locus_tag="SPC15C_0001"

```

CDS
<1..134
/gene="dexB"
/locus_tag="SPC15C_0001"
/codon_start=3
/transl_table=1
/protein_id="CA133379.1"
/db_xref="GI:8643073"
/translation="SNESQDLTVEGKYSVLIENTLAQEVFEKQILVPADACVELL"
293..445
/gene="al1B"
/locus_tag="SPC15C_0002"
/pseudo
293..445
/gene="al1B"
/locus_tag="SPC15C_0002"
/pseudo

gene
/transl_table=1
/product="putative oligopeptide-binding protein Al1B
(pseudogene)"
/db_xref="PSEUDO:CA133380.1"
complement(462..981)
/gene="tnp"
/locus_tag="SPC15C_0003"
/pseudo
complement(join(462..521,520..981))
/gene="tnp"
/locus_tag="SPC15C_0003"
/notes="member of homology group 0004 74"
pseudo
/codon_start=1
/transl_table=1
/product="putative IS630-Spn1 transposase fragment"
1204..2649
/gene="wzg"
/locus_tag="SPC15C_0004"
1204..2649
/gene="wzg"
/locus_tag="SPC15C_0004"
/notes="member of homology group 0000 90"
/codon_start=1
/transl_table=1
/product="Integral membrane regulatory protein Wzg"
/protein_id="CA133382.1"
/db_xref="GI:86643074"
/translation="MSRRPKSSSQKVRKSVNIVLTITVLLNCFLLFLFKYNIATF
RINLVVTVLIVLVALVGLLITIKYAKKPTFLVPSIVLSVSLFAVOQFVGLNR
LNATSVSYSTISVAVLADSDIENVYQITSTYAPGTDNENIQKLDIKSSONTDT
VOQSSYLAAYKSLIAGETKALVINSVENITIESEYEDYASKIKKITYGKTKVEAF
KTSKQNFPIVYSIDITYGPISSVSRSRVNIMTVNRDRTKTLITTPRAVYVADIAG
KNNQKDKIHAGIYGVDSIHTEMLYGVGINVYVRLNFTSLKMLDGLGVHDNQ
EESALHGKHFPGVGNVHLDSEQALGFVERSLADGDRGNQOKVIVAILQKLTST
EALPKYSTIINSIQDSIQTNMPLFETMINLVNAQLSESGNRYKNSQDLKGTGRDLPXY
AMPKSNLYLEIDSSLAIVKKAIDVMEGR"
1204..1320
/gene="wzg"
/locus_tag="SPC15C_0004"
/notes="Signal peptide predicted for SPC0720 by SignalP 2.0
HMM (Signal peptide probability 0.997 with cleavage site
probability 0.540 between residues 39 and 40"
join(1252..1311,1339..1407,1426..1494)
/gene="wzg"
/locus_tag="SPC15C_0004"
/notes="3 probable Transmembrane helices predicted for
SPC0720 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
1411..1755
/gene="wzg"
/locus_tag="SPC15C_0004"
/notes="HMMpfam hit to PFO2916, DNA polymerase processivity
factor, score 5.4e-63"
1933..2376
/misc_feature
/locus_tag="SPC15C_0004"
/gene="wzg"

```

gene	/note="HMMpfam hit to PF03816, Cell envelope-related transcriptional attenuator, score 5.2e-70"
	2651. .3382
	/gene="wzh"
CDS	/locus_tag="SPC15C_0005"
	2651. .3382
	/gene="wzh"
	/locus_tag="SPC15C_0005"
	/EC_number="3.1.3.48"
	/note="member of homology group 0001 90"
	/codon_start=1
	/transl_table=1
	/product="protein-tyrosine phosphatase wzh"
	/protein_id="CA133383.1"
	/db_xref="GI:68643075"
	/translation="IMDISHIVFDVDDGPKRSRSKALLAESRGVRIIVSTSHR
	KGMETPEEKIAENFYDIHSAIKVASDVIYAGATVYTPVDLDEKERRIPLTDS
	RYALIESMTPTPYDHSALSKIMLGITPVIHTRYPDALENNERGAVELDMCGYT
	QVNSHVLKPLFGERYFMKRAQYFLERDLVHVASDMDNLSRPYMOQAYDIIA
	KYGAKKAKEFFVFNPKRIIMDOLI"
	2963. .3277
misc_feature	/gene="wzh"
	/locus_tag="SPC15C_0005"
	/note="HMMpfam hit to PF02811, PHP, C-terminal, score
	3.5e-13"
	3391. .4083
	/gene="wzd"
gene	/locus_tag="SPC15C_0006"
	3391. .4083
CDS	/gene="wzd"
	/locus_tag="SPC15C_0006"
	/note="member of homology group 0002 90"
	/codon_start=1
	/transl_table=1
	/product="capsular polysaccharide biosynthesis protein
	wzd"
	/protein_id="CA133384.1"
	/db_xref="GI:68643076"
	/translation="MKEHNTLEIDILQPLALMKRKLIVLLVAITTSVAFASTVY
	KPESTSTRIVVNRNOEKSLGNODLGTYVVKRYREILISQDLAEVILSDKLD
	ITPKCLAKIVYTPVDRIYYSVNVNRVPEEARSINARSLEVAQOKIISITPVSPT
	TLEERKPAISPSVPNIKRNILIGFLAGVGIVALLLELDTHVKRPEDITLQMT
	LLGVNPLSKLK"
	3409. .3825
misc_feature	/gene="wzd"
	/locus_tag="SPC15C_0006"
	/note="HMMpfam hit to PF02706, Lipopolysaccharide
	biosynthesis, score 8.8e-54"
	join(3457. .3522,3925. .3993)
	/gene="wzd"
	/locus_tag="SPC15C_0006"
	/note="2 probable Transmembrane helices predicted for
	SPC0722 by TMHMM2.0 at aa 23-44 and 179-201"
	3931. .4008
misc_feature	/gene="wzd"
	/locus_tag="SPC15C_0006"
	/note="PS00217 Sugar transport proteins signature 2."
	4093. .4776
gene	/gene="wze"
	/locus_tag="SPC15C_0007"
	4093. .4776
CDS	/gene="wze"
	/locus_tag="SPC15C_0007"
	/EC_number="2.7.1.112"
	/note="member of homology group 0003 90"
	/codon_start=1
	/transl_table=1
	/product="tyrosine-protein kinase wze"
	/protein_id="CA133385.1"
	/db_xref="GI:68643077"
	/translation="MPTLEIAQKLEPTKAAEYNNALCTNIQSGDKVIVSYVNV
	PEGICNTNENFVAVSAGYKTLTIDDTNYSVSGFEKSRKTEITELPSTADL
	SHGLCTNTNENFVAVSAGYKTLTIDDTNYSVSGFEKSRKTEITELPSTADL

gene
 DAA110KCDASIIIVTATGEVNRKDVOKAQOLEQNKPLGLVILNKPDVQHKYGSY
 GDYGVGKK"
 4792..6159
 /gene="wcha"
 /locus_tag="SPC15C_0008"
 4792..6159
 /gene="wcha"
 /locus_tag="SPC15C_0008"
 /note="member of homology group 0005 66"
 /codon_start=1
 /transl_table=11
 /product="undecaprenylphosphate glucosyltransferase
 wcha (initial sugar transferase)"
 /protein_id="CA133386.1"
 /db_xref="GI:68643078"
 /translation="MDEKGLKIFLAUVQSIIVILVFLSFVRETELEKSSWILYL
 HFPEVSSYGNPFKRGHLVEFNSTRIRYFFPAIALSVINFAEFSISRMVVF
 LTLFGLSLYLINLPLVKRYMKRVFENLNSKKILILITVKMKEKILDLSDLSWML

Query Match 12.2% Score 851.8; DB 1; Length 18626;
 Best Local Similarity 62.2%; Pred. No. 3.4e-77;
 Matches 1447; Conservative 0; Mismatches 857; Indels 21; Gaps 6;

QY 16 TGGCATTATTTGATATGATAGACAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCA 75
 |||||
 Db 4820 TGGCAGATTACAGAGATTATTGTCATTTATGTTATTTCTTAGCTTTGTTAGAG 4879
 QY 76 ATGCTGATTTAATCGTTCTGGAATTTTATCATATGATGTCATTTATTTGCAATTT 135
 |||||
 Db 4880 AGACAGAACTTGAACGTTCTCGATGGTATGATCTATACCTTCACCTTTTGTATTTCT 4939
 QY 136 TTAATCTCGATCCGAGTTGAATTTGATATAGATTAATCTGATAGCTTTGAAAAA 195
 |||||
 Db 4940 ATTTTATCTCTATGATATATTTTAAAAAGAGGCACTTATGTTAGTTAATAGTA 4999
 QY 196 CATTATCTATGATATATATTTGCAATTTTCTTAAGGAGATCAATTTTGTGAGA 255
 |||||
 Db 5000 CTATATATATATTTTCTTTCGCAATAGCTATATAGTATTAACCTTTTATACGG 5059
 QY 256 ATAAATTTGCACTTTCAGAGCGTGCGCTGATATTTCACTAATAAATCTCGTTTGG 315
 |||||
 Db 5060 AACGTTAGATCTCTAGAAAGAGAAATGATATCTTAACTTTAGAGAAATATCT 5119
 QY 316 TATATCTATTAACGTAATTTATTAAGCAGTTAAGATAGCTTCTATTTCCAGATCT 375
 |||||
 Db 5120 TATATCTGTTAAATTTCTTATGTAAGAAATATGAAAGCATGTTTTTTAATCTAAAA 5179
 QY 376 ATCAAAAAAGACATTTCTAATTTCAACGCTGAACGATGGGAAATATGCAATTTTAT 435
 |||||
 Db 5180 ATAGCAAGAAATTTTACTGTTAACAGTAAGAAATATGAAAAAGTTCTTGATAAAT 5239
 QY 436 TTGATCATCATTAACAATTCAAAAAATCTTGTGCAATGGTAGTTTAGAGTACAGAA 495
 |||||
 Db 5240 TGTAGATCTGATGAACCTTTCATGGAATGATAGAGTAAGTGTGATTAATCTG 5299
 QY 496 TAGATTAATTAATTTATCATTTACCGCTCTATTTTCTGTGGAAGAAAGCTATAGTTT 555
 |||||
 Db 5300 ATTTTCAACATGATTAATATCTGTAATGAAAA-----GAAAAAATATTAATCAATTTG 5353
 QY 556 CAACAAGGAAGTGGTGAACGCTCTTATTAATCTTACCAAGTGGTTTATAGAGTAA 615
 |||||
 Db 5354 CAACGATGAAGTTGTGAAGAGGTGTTTGTCAATCTTCCAGAGAGAGCTACATATTTG 5413
 QY 616 AGCAATTCGTTTCAGATTTGAGTGTGATGATGATAGTAAGCGTTGATATTAATTCAT 675
 |||||
 Db 5414 GAGAAATTAATCTAGGTTTGAACAATGGGATAGATGTAACGTAAATCTTAAACAT 5473
 QY 676 TCGGTTTACGCGTTGAAAAACAAAAATCCAACTGCTAGGTGACATAGCATTTGTA 735
 |||||
 Db 5474 TTGATGAAGATTTGGTGGCAGATTAACAATTCATGAGATGATGATGATGATGATGCA 5533
 QY 736 CTTTTCACAAATTTTATTAAGCTAGTCATATCATGATGAAGCACTTTGATATAC 795
 |||||

Db 5534 CTTTCTCTAACAATTTTATTAATAAATAGTCATGATTTCAAGAGAAATTCGTATATTT 5593
 QY 796 TCGAGCGGTAGTACGCGGTATATTTATTTGGTATAGTTTCTATTTGTTAGTTCCAAATTA 855
 |||||
 Db 5594 GTGTGACCATTTATGCGCTTCTTTTGTCTATAGCTAGCTAGTTTATGTTCAATTTGA 5653
 QY 856 TTGCTAGAGATGTGACCGGCTATTTTGTCTAGAAACGAGTTGACAGAAATGACGCA 915
 |||||
 Db 5654 TTGCTAAAGATGGGACGACGCTATTTTGTCTCAAACTCGTATAGGAAAAATGCTGCAC 5713
 QY 916 TATTTACATTTCTACAGTTTGCATTCGATGATGTGATGCTGAGAGACGCAAAAAAGCT 975
 |||||
 Db 5714 ATTTTACCTTTTATTAATTCCTGTCGATGAGGATCGATGCTGAAGCTATCAAAAGACGT 5773
 QY 976 TGGTCACCAAAACCAATGCAAGGCGGTATGTTTAAATGAGAAAAACGATCTCTAG 1035
 |||||
 Db 5774 TGATGATCAAAATATCATGAGAGGTGATGTTT-----AAGATGCAATATGATCTCTG 5828
 QY 1036 AATTACTCCAATTGACATTTTCATACCGAAAAACAGTTTATAGACGTTTACCAAGTTT 1095
 |||||
 Db 5829 TGTTCAGAAATATGCTGCTTTATTCG-TAAACAGTTTGGATGAAATGCGCAGTTT 5887
 QY 1096 ATAAATGTTTATTTGCGGATATGATCTAGTTGTACACGTCACCTACAGTTATGAT 1155
 |||||
 Db 5888 GGAATGCTTTATATGAGATATGATGATGTTGGTGGGAAACGTCACCTACAGTATACAGAT 5947
 QY 1156 TTGAAAAATATATACCTCCGTGCAAAAGAGATGATGATTTTAAACGAGGATTAACAGGTC 1215
 |||||
 Db 5948 ATATGATGATATCTCCAGAACAGAAACGTGACCTCACTTAACTCGTATTAACAGTT 6007
 QY 1216 TCGGACAGTTAAGTGTGCTAGTATATCAACAGCTTCGACGATGTTGCTGTTGACT 1275
 |||||
 Db 6008 TATGCGAGTTATGCGCCGTATGTAATTAACGATTTGACGATGTTGTAATTAATGATG 6067
 QY 1276 TAGCATACATGATATATGAGCTATCTGTCGATATATTAATTTTATTAAGACAGTGA 1335
 |||||
 Db 6068 TGGCTTATATGATATATTTGACCAATCTGAAAGATATTTGCTTAAACTGTGTA 6127
 QY 1336 AAGTTGATTTGTCAGAGAGGAAATGATTAAGAAATATGAAAGTTTGTGTCGTT 1395
 |||||
 Db 6128 AAGTTGATTTATTAAGATGAGAGCAAGTAA--GGAATGAAGATATGCTTATGTTGAT 6184
 QY 1396 CTTCAAGGAGACATTTGACTCATCTTGTATTTGTTAAACGTTTGAAGAGAAAGAAC 1455
 |||||
 Db 6185 CTAGTGAGGCAATTTAATCTCATTTGTATTTATTAACCTTCTGGAATGATTAAGAA 6244
 QY 1456 GTTTTGGTAACTTTGATTAAGAGAGATCAAGAACTTTTGAAGATGAAGAAAAATGT 1515
 |||||
 Db 6245 GATTTGGTTACTTTTATATGATTAAGCGGACGCTCGAGATGTAATCTGAGAGAAATGTAAT 6304
 QY 1516 ATCATGTTACTTTCCAAACAAATCGCAATCTCATTAATTTAGTGAATAAATCTTCTTAG 1575
 |||||
 Db 6305 ATCCCTGTTATTTCCAAACCAATCTGTAATTAATAAATAAATAAATAAATAAATAAATA 6364
 QY 1576 CTTTCAAAATTTTAAACGATGAGAAACCAAGTGTATTAATTTCACTGTGCGCGCTGTG 1635
 |||||
 Db 6365 CTAATAAATTTTAAAAAGAGAGACCAAGTTTGAATTTCTTCTGTGCGCGCTATGCG 6424
 QY 1636 CTGTCCCTCTTTTATATGTTGGAAGTATTTTGGATCAAGAGACGTATATATAGAGTGT 1695
 |||||
 Db 6425 CTGTTCCTTTCTTTTATATGTTGGAAGTATTTTGGATCAAGAGACGTATATATAGAGTGT 6484
 QY 1696 TTGATCGAGTTAATTAATCTACATTAACGTGAAAACTAGTTTATCCGCTAACAGATATTT 1755
 |||||
 Db 6485 TTGATTAAGATTTGATCTTCAACACTGACAGAAAACTGTGTTTCCAGTAAACGATAGT 6544
 QY 1756 TTAATGTCAGTGGAGAAATGAAGAGGTATATCTTAATCTATTAACCTTGGGAGTA 1815
 |||||
 Db 6545 TTATTTGATCAATGAGAAAGAAATGAAGAAATGTAATTAACCAAGCAATTAACCTAGAGAGGA 6604
 QY 1816 TTTTATTAATGATTTTGTAAACGTAGAACTCATGAACACAGTTTATGATGATTAAT 1875
 |||||
 Db 6605 TATTTCTATATTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATTA 6664

QY	1876	GAGATTGATTTATTTGAAAAAATAATGGAAGTATACGACGAATATTTATCGAAACGG	1935
Db	6665	AGAGGTGATTATCTAAAAAAGAAAATCTTATTCAGATGAGTATTTATTCAAATGCG	6724
QY	1936	ATATTTCTGACTATATTCGAGATATTCGACGATATAAAAATTTCTCAGTTACAAAGAAAT	1995
Db	6725	ATATTTCCAGTTATATTTCCGAATATTTGGAGTGGGAAAAAGATATTTCTTATGAAAAAAT	6784
QY	1996	GGAACAAATATTTATTAACAATCAGAAGTAGTTATTTGCCAGGAGCCCGCTACTTTAT	2055
Db	6785	GAATCAATTTGATTAAAGAAACAGATATTTATCAATTCTCATGTGACACGAGCACTTTAT	6844
QY	2056	GAATTCATTATCCAAAGAAAAAACAATATATTTGTTCTCAGACAAAAAAGTAGTGTA	2115
Db	6845	GGGGGTTATTTGCTAAGAGAAAAAGTTCCGATAGTTGTTCCGAGGCAAAAAAATTTGGAGA	6904
QY	2116	ACATGTAAATGATCATCAATGAGTTTGTAAGAAATTTTACACATTAATATATTTT	2175
Db	6905	GCATGTGATGACACCTTCAGCTTCATTTGCTAATCTGACGAAGAAATATTCAAATTTTAT	6964
QY	2176	ATTTATAAAAAATATAGATGATTTGTTGAAAAAATTAATGAGTTCTTAGCAACAT--	2233
Db	6965	AGCTCAATGATGATTTTCCGACTTATATTTAGTATACGTGATTTTAAAGATTAATATTT	7024
QY	2234	--AACTTTACATCAATATATATTTTTTTTGTGAAAGATTAAAC--AAATAGTTGAAA	2289
Db	7025	TGGAATTTGTTGAACATGAAAGATTTAATCTACATTTTATGCGGAAATTAATTAAC	7084
QY	2290	ATTTATATGAGATCAAGAAAATGATTAATAAAAAGATGCANATTT	2334
Db	7085	ATTTGAGGAATTAAGTAGAAGAAAATTTAGATGAGACCTTCAT	7129

CR9311657	RESULT 15
LOCUS	CR9311657 16926 bp DNA linear BCT 30-JUN-2005
DEFINITION	Streptococcus pneumoniae strain 34356 (serotype 11f).
ACCESSION	CR9311657
VERSION	CR9311657.1 GI:68642865
KEYWORDS	
SOURCE	
ORGANISM	Streptococcus pneumoniae Streptococcus pneumoniae Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE	1 (bases 1 to 16926) Bentley, S.D., Aanensen, D., Mavroidi, A., Saunders, D., Rabinovich, E., Collins, M., Donaghe, K., Harris, D., Kallott, M.S., Murphy, L., Quail, M.A., Samad, G., Skovsted, I.C., Barrall, B.G., Reeves, P., Parkhill, J. and Spratt, B.G. Genetic analysis of the capsular biosynthetic locus from all 90 serotypes of Streptococcus pneumoniae Unpublished
JOURNAL	2 (bases 1 to 16926) Bentley, S.D.
REFERENCE	Direct Submission Submitted (09-DEC-2004) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: sds@sanger.ac.uk
COMMENT	NOTE: This sequence was generated from a PCR product representing the region from dexB to alyA and is not necessarily responsible for the expressed capsule serotype. For a detailed description of how CDS products were predicted see the associated publication.
FEATURES	Location/Qualifiers
source	1..16926 /organism="Streptococcus pneumoniae" /mol_type="genomic DNA" /strain="34356" /db_xref="taxon:1313" <1..137 /gene="dexB" /locus_tag="SPC11F_0001" <1..137
gene	
CDS	

```

gene
    complement(268. .1679)
    /gene="tnp"
    /locus_tag="SPC11F_0002"
    /pseudo
    complement(join(268. .963,.963. .1679))
    /gene="tnp"
    /locus_tag="SPC11F_0002"
    /note="member of homology group 0017 25"
    /pseudo
    /codon_start=1
    /transl_table=11
    /product="putative IS1202 transposase"
    /db_xref="PSEUDO:CA133196.1"
    1848. .3302
    /gene="wzg"
    /locus_tag="SPC11F_0003"
    1848. .3302
    /gene="wzg"
    /locus_tag="SPC11F_0003"
    /gene="wzg"
    /locus_tag="SPC11F_0003"
    /note="member of homology group 0000 90"
    /codon_start=1
    /transl_table=11
    /product="integral membrane regulatory protein Wzg"
    /protein_id="CA133197.1"
    /db_xref="GI:68642867"
    /translation="MLMSRRFKKSGSQKRVSNVILITLYLLVCFLLFLPIFKNI
    LAFRIALVIVLVLVALVGLILIIYKKEKFTIPLLSILVSSVSLPTVQSPVL
    LTRFATNVNSEYSISVAVLADSEIENVTLQTSVTAFTGDNENIIOKLADISQML
    DLTVNASSSYLAAYKSLIAGDTKAIIVNSVFNIISESPYAKIKIYKPTKKV
    EAPKTSQNSFRITYSGIDITGPITSSVSRSDNVIIMTNVDTKKIILLTPRDAVYDI
    ADGGNNQDCKLTHAGTIGVDSISHLENTLGVNDINYVRLNFTSPFLKIDILGVDVNH
    NDQESALHKGHFEPVGNVHLDSEQLGFEVERYSIADGDRGRNOQKVALLOKL
    TSTELAKVNGTILNLSQSIQTNPLETMDIVNTQLESKGKRVNSQDLKGTGRMDL
    PSYAMPDNLYLEIDDSLAVVKAALQDVMEGR"
    1848. .1973
    /gene="wzg"
    /locus_tag="SPC11F_0003"
    /note="Signal peptide predicted for SPC0523 by SignalP 2.0.
    HMM (Signal peptide probability 0.991) with cleavage site
    probability 0.537 between residues 42 and 43"
    join(1905. .1964,1992. .2060,2079. .2147)
    /gene="wzg"
    /locus_tag="SPC11F_0003"
    /note="33 probable transmembrane helices predicted for
    SPC0523 by TMHMM2.0 at aa 20-39, 49-71 and 78-100"
    2064. .2408
    /gene="wzg"
    /locus_tag="SPC11F_0003"
    /note="HMMpfam hit to PF02916, DNA polymerase processivity
    factor, score 1e-62"
    2586. .3029
    /gene="wzg"
    /locus_tag="SPC11F_0003"
    /note="HMMpfam hit to PF03816, Cell envelope-related
    transcriptional attenuator, score 7.7e-70"
    3304. .4035
    /gene="wzh"
    /locus_tag="SPC11F_0004"
    3304. .4035
    /gene="wzh"
    /locus_tag="SPC11F_0004"
    /BC_number="3.1.3.48"
    /note="member of homology group 0001 90"
    /codon_start=1
    /transl_table=11

```

	Query Match	Best Local Similarity	Matches 1434; Conservative	Score 851.6; DB 1; Length 16926;	Signal peptide predicted for SPC0527 by SignalP 2.0 (signal peptide probability 0.844) with cleavage site
QY	14	ATTGGCATTATTGATATGATAGTACAGTGGCAATTTCTGCAATCTTAAACGATATATAC	73		
DB	5476	ATTGGCCATATCCAGAGTTTCTTGTTATTTATGACTTATCTACTTAGTGTCTGGAG	5535		
QY	74	AAATGCTGATTTAAATCGTTCCTGAAATTTTATCATATAGATGTTTATTTTGCATT	133		
DB	5536	AGAAACGGAATGTTTCAACAAACAGCTATGCACTTATATCTCATATTATTTGCTT	5595		
QY	134	TTTTATATCTCGTATGCGCAGTTGAATTTAGTATATAGAGTATCTGATAGATTGAAA	193		
DB	5596	TTTATATCAGATGATTATGACAGAGATTTCTTTAAAGAGATATTATTAATGAACCTTGCCA	5655		
QY	194	AAACATTTAACTATAGATAATATTTGCAATTTTCTTACGCGAGTATCACTTTTGTGGA	253		
DB	5656	GACATTTGAATATATATCTTATTTCTTGACACTAGGATTTAGTATTTCTAATTTTCTTAGA	5715		
QY	254	GAATTAATTTGCACTTTCAGACGCTGATGTCCTGTTATTTTCACTAATAAATCTGTTTT	313		
DB	5716	GGATCGATTTTATGATTTTCCAGACGAGGACGATTTACTTCTCCACATTAATATGCTCTT	5775		
QY	314	GGTATACCTATTTTAACGTAATTTTAAACAGTTTAAAGATAGCTTCTATTTTCGACAA	373		
DB	5776	AGTCTATGTCCTAAACCTATTTATTCAGAGTGTTATGGAACCGCATTTATCCAACTTTAA	5835		
QY	374	CTATCAAAAAAAGACATTTCTATTTCAACCGCTGAACGATGGGAAAATATGCAAGTTT	433		
DB	5836	AGGAAGTAAAGAAATTTCTCTACTTACAGCACTTCTCGTGTGGAAGAGTACTGCATAA	5895		
QY	434	ATTGATATCATAAACAATTCMAAAAAATCTGTTGCAATTTGTTAGTTAGTACAGA	493		
DB	5896	ATTATATGAAATCAAAATGAGGTGTGGGAGTTGTTAGCCGTGAGTGTCTTAATATAAC	5955		
QY	494	AATAGATAAAATTAATTTATCATTTACCGCTCTATTTATTCGTGGAAGAACTATTAAGTT	553		
DB	5956	AGATTTTCAGCATGATTTGTTTAAAGTAGT-----AGCAGAGGGGAGATAGTAAACCT	6009		
QY	554	TTTCAACAGGAGAGTGTGACACGCTCTTATTAATCTAACAAAGGAGTTTTTAAGC	613		
DB	6010	TGGACATCATGAGTGTGTGATGAAATCTTTTAAACATCTTCCAAAGTGAAGAAATACATAT	6069		
QY	614	AAAGCAATTCGTTTCAGATTTTGAATTTGTTAGTATGATATGTAAGGTTGATTTTAATTC	673		
DB	6070	TGAGAGCTGTGTCTCTCAGTTTGAACATGGGAATTTGATGTAAACGTCAATTAATAC	6129		
QY	674	ATTGCGTTTATCTGCGTTGAAAAACAAAAATCCAACTGCTAGTGAACATAGCATTTGT	733		
DB	6130	TTTTGATCGTATTTTGGCACTTAAACAGAAATCCGTGATAGTGGCAGATTAATATGTGT	6189		
QY	734	AACTTTTCCAAATTTTATTAAGCTGTGATATCAGATGAAACGATTTTGGTAT	793		
DB	6190	GACTTTTCTACAACTTTTATTAAGACTAGTCAATGTAATTTGCTTAACGAGATTTATATAT	6249		

QY 794 ACTCGAGCGGTAGTGGGTTAATTATTTGTGTATAGTTTCTAATTTGTTAGTTCCAA 853
 Db 6250 TGTGGTGTGATGTGTAGGGCTGATCTATGCGGTCTAGTCAGTGTGTACTGGTCTTT 6309
 QY 854 TATTCGTAGAGATGTGTGACCGGCTATTTTGTCTCAAAAAGATTGACAGATGTGAG 913
 Db 6310 GATTCGAAAAGATGGGGGCTCTGCTATTTTGTCTCAGACGCGTATTTGAAAAATGTAG 63169
 QY 914 CATATTTCATCTCACAAGTTTCGATCGATGTATGTGTGATGCGAGAGCGCAAAAAAG 973
 Db 6370 ACAATTTTACTTTTAAAGTTTGCTCCATGTGTGTAGATGTGAGAGCAAAAAAGAA 6429
 QY 974 CTGTGTCAGCAAAAACAAGTGCAGAGGGTGTATGTTTAAATGG3AAAAAGATCTT 1033
 Db 6430 ACTCATGGAACAAAATACCATGCA-----GGTGGATGTTTAAAGTGTGATGTGCCA 6484
 QY 1034 AGAATTTACTCCAAATGTGACATTTTCATACGCAAAAACAAGTTTACAGATTTACCAAGT 1093
 Db 6485 CGTATCACGAAAATTTGGTCTGTTTAAATACG-GAAGACTAGCTTGAGACGAGCTGCCACA 6543
 QY 1094 TTTATATGTTTAAATTTGGCGATATGAGTCTAGTTGTGACGCTCCACCTACAGTTGTGA 1153
 Db 6544 TTATATATGTTCTAAAGGAGATATGAGTTGTGTGACACGCGCACCAACAGTGGACGA 6603
 QY 1154 ATTTGAAAAATATATCTCTGTGCAAAAAGACGATTTGATTTTAAACAGGATTTACAG 1213
 Db 6604 GTATGAACCTATACCCCAACAAAACGTCGTTAAGTTTAAACCTGTATTAACAG 6663
 QY 1214 TCTCTGGCAGGTTAGTGTGTGTATATACAGACTTGACAGACGTAAGTTGCGTTGGA 1273
 Db 6664 CTTATGGCAGGTAGTGTGACGAGAGTTAAGAAATTTGATGAGTTGTCAAAATTGA 6723
 QY 1274 CTTAGCATACATTTGATATTTGACTATCTGTGACAGATTTAAATTTTAAAGACAT 1333
 Db 6724 TGTAGCCTATATAGACGATTTGACAAATTTGAAAAGATTTGAAATTTTATGAACAGT 6783
 QY 1334 GAAAGTTGTATTTGTGTGAGAGGAGAAATTAAGTAAAGTATATGAAAGTTTGTGTCG 1393
 Db 6784 TAAAGTTGTATTTGTGTGAGAGATGTGAGCAAGT--AGAAATGAAAATATGTTTGTG 6840
 QY 1394 TTTCTGAGGGGACATTTGACTCACTTGTATTTGTAAAAACGTTTGAAGAGAAAGA 1453
 Db 6841 ATCTAGTGTGAGGTGATCTGACTGATTTATCTTTTAAACCTTTCTGAAAAGATAAGA 6900
 QY 1454 ACGTTTTGGGTAACTTTGATTAAGAGATGTCAAGAGTCTTTGAAAGATGAAAAAT 1513
 Db 6901 AAGATTTTGGGTAACTTTGATTAAGGCGGACGCTCGAGGTACTCTCTGAGGAAAGTGT 6960
 QY 1514 GTATCCATGTTACTTTCCAACAATTCGAATCTCATTAATTTAGTGAATACTTTCTT 1573
 Db 6961 ATATCCCTGTATTTATCCAACCAATCTGATATATTAATAATAATAATAATACTTTCT 7020
 QY 1574 AGCTTTCAAAATTTTACGTGTGTGAAAAACAGATGTATATTTTCACTGTGTGCGCGT 1633
 Db 7021 CGCTATATAAATTTTAAAAAAGAGAGCCGATTTGATTTCTTCTGTGTGCGCGCTAT 7080
 QY 1634 TGTGTGTCCTCTTTTATCATCGAAAACTATTTGAGCAAAAGACATTTATTTGAAGT 1693
 Db 7081 CGCTGTCTCTTTCTTTATATTTGGAAGTTATTTGATCAAAAGACAGTATATATAGAAGT 7140
 QY 1694 ATTTGATCGATTAATTAATCTACATTAAGTGAATACTAGTTTATCCGTAACAGATAT 1753
 Db 7141 GTTTGATAGAAATTAATGCTCCACACATTAACGGAATACTTTTATCCAGTAACGATTA 7200
 QY 1754 TTTTATTTGTCAGTGGGAAAGATGAAGAGTATATCTTAATCTAATTTAATTGCGGAG 1813
 Db 7201 GTTTATTTGTTCAATGGGAAAGATGAAGAAATTTATCCAAAGCAATTAATCTAGGAGG 7260
 QY 1814 TATTTTAAATGATTTTGTGAACAGTAGAACTCATGAACAAGTTTAAATGATGATA 1873
 Db 7261 AATATCTGATGATTTTGTGTTACTGTAGGACACATGAACAATTTGATGTTGTGATA 7320

QY 1874 AAAGATTTGATTTATTTGAAAAAATGAAAGTATTAACCGAGAAATATTTTCAACA 1933
 Db 7321 AAAGAAAGTTGATCGTTTGAAGAGAAATCTAATTTCAAGAGAAATATTTATCAACA 7380
 QY 1934 GGATATTTGATATATTTCCAGATATTTGCAAGTATTAATAAATTTCTCAGTTACAAAGA 1993
 Db 7381 GGATATTTCTAATTTATGTCCCAAAATTTTGTAAATGGGAAAAATATATCTATGAAAA 7440
 QY 1994 ATGGAACAATATATTAACAATCAAGAGTATTTTGTGACGAGAGGCGCGTACTTTT 2053
 Db 7441 ATGAATCAATTTATGAGGAATCTGATACAAATTTGTACCCATGTGTGTCCAGTACTTTT 7500
 QY 2054 ATGAATCTATTTCCAAAGAAAAAACAATTTATCTTTCTTACAGCAAAAAAGTATGTT 2113
 Db 7501 ATGGCAATTTATGCTAAAGAAAAAATCTCAATATGTTGTCCAGAGCAAAAGAAATTTGG 7560
 QY 2114 GAACATGTAATGATCATCAAGTAGAGTTGTGAAGAGATTTTACAAAGTATATATAT 2173
 Db 7561 GACATGTGAATGATCATCAAGTGAATTTTGTCTTAAATATCCCTGAAAGGTATGAA 7620
 QY 2174 TTAATTTAAGAAAAATATAGATATTTGTTGAAAAAATTTATGAAGTTTCTAAGCAACT 2233
 Db 7621 TTAATCTGATTTAAGATATTAAGCAATTAACAGTATAGTTTAAAGCAAGAAAAAAGAACT 7680
 QY 2234 AACTTTACATCAATATATATTTTGTGTGAAGATTTAAACAATAGTTGAAAAATTT 2293
 Db 7681 AATTTTACTCTAATATATCAAAATTTTCTTAAGCAGTATCAAAAAATTAATGAAGATTTA 7740
 QY 2294 AATGAGATCAAGAAAAATGAATTAATAAAGATGCAATTTGATTAAT 2341
 Db 7741 ATAGTATTAAGATGAGATTTGTTCTTCTTATGCAAGATGAATTAAGAAAT 7788

Search completed: December 25, 2005, 11:35:42
 Job time : 22541 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2005, 18:36:42 ; Search time 2504 Seconds
(without alignments)
18610.036 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgcaacgaatgca.....gcaaacgcaaaagaagctt 6992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001s: *
5: geneseqn2001bs: *
6: geneseqn2002as: *
7: geneseqn2002bs: *
8: geneseqn2003as: *
9: geneseqn2003bs: *
10: geneseqn2003cs: *
11: geneseqn2003ds: *
12: geneseqn2004as: *
13: geneseqn2004bs: *
14: geneseqn2005s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6992	100.0	6992	3	Aaz60930 Nucleotide
2	1195.8	17.1	26281	3	Aaz60929 Nucleotide
3	955.6	13.7	17468	6	Abao1441 Streptoc
4	889	12.7	16032	13	Adsl3200 S. thermo
5	798.8	11.4	110000	6	ABN71527_11
6	727.8	10.4	6865	6	ABK90550 Beta1,3-G
7	726.8	10.4	17276	8	ACA64723 Streptoc
8	726.8	10.4	17276	10	ADR43363 Streptoc
9	726.8	10.4	17276	14	AEA03034 Streptoc
10	725.2	10.4	95596	13	ADV87741 Streptoc
11	725.2	10.4	95596	13	ADV78994 Streptoc
12	725.2	10.4	110000	13	ADV81804_12
13	721.4	10.3	25020	12	ADO40235 S. agalac
14	649.2	9.3	2226	11	ADM79774
15	428.4	6.1	6850	4	AaH43197
16	423.2	6.1	1368	4	AAS5950 Streptoc
17	415.8	5.9	1368	13	ADR44817 Streptoc
18	415.8	5.9	1443	13	ADR93668 Novel S.
19	415.8	5.9	1443	14	AEA57538 Streptoc

20	383.2	5.5	18274	6	ABA01446	Abao1446 Streptoc
21	376.2	5.4	19966	3	Aaz230355	Aaz230355 Nucleotide
22	376.2	5.4	19966	3	Aaz45258	Aaz45258 DNA encod
23	376	5.4	28170	6	ABA01447	Abao1447 Streptoc
24	373	5.3	15962	6	ABA01445	Abao1445 Streptoc
25	371.6	5.3	20555	6	ABA01443	Abao1443 Streptoc
26	371.4	5.3	18488	6	ABA01444	Abao1444 Streptoc
27	371.4	5.3	19738	6	ABA01446	Abao1446 Streptoc
28	366	5.2	25243	6	ABA01442	Abao1442 Streptoc
29	364.8	5.2	22157	6	ABA01448	Abao1448 Streptoc
30	347.2	5.0	14142	6	ABA01435	Abao1435 Streptoc
31	346	4.9	12786	6	ABA01433	Abao1433 Streptoc
32	332	4.7	14460	6	ABA01440	Abao1440 Streptoc
33	324.8	4.6	1389	13	ADV84472	Adv84472 Streptoc
34	320.2	4.6	1386	6	ABN70387	Abn70387 Streptoc
35	320.2	4.6	1386	6	ABN67429	Abn67429 Streptoc
36	314.8	4.5	15347	6	ABA01439	Abao1439 Streptoc
37	301.4	4.3	8056	8	ABZ10246	Abz10246 Haematopo
38	298.8	4.3	8056	8	ABZ10246	Abz10246 Haematopo
39	298.4	4.3	969	13	ADV84476	Adv84476 Streptoc
40	291.6	4.2	16110	6	ABA01437	Abao1437 Streptoc
41	283.8	4.1	1002	10	ADC50017	Adc50017 N-acetyl
42	259.6	3.7	450	13	ADV84473	Adv84473 Streptoc
43	255	3.6	447	6	ABN67430	Abn67430 Streptoc
44	230.2	3.3	12494	6	ABA01438	Abao1438 Streptoc
45	222.2	3.2	8056	8	ABZ10100	Abz10100 Haematopo

ALIGNMENTS

RESULT 1	
ID Aaz60930	standard; DNA; 6992 BP.
XX Aaz60930;	
XX 30-MAY-2000 (first entry)	
XX Nucleotide sequence of a capsular gene cluster of S. suis serotype 1.	
XX Capsular gene cluster; serotype 1; polysaccharide biosynthesis;	
KW capsular component; antigen; regulation; chain length determination;	
KW complement-mediated opsonophagocytosis; serotype-specific detection;	
KW antigen; vaccine; Streptococcal disease; CpsA; CpsF; CpsH;	
XX CpsI; CpsJ; CpsK; glycosyltransferase; Cp polymerase; se.	
XX Streptococcus suis.	
XX Key	Location/Qualifiers
FT CDS	2..1367
FT	/*tag= a
FT	/note= "CpsA; encodes AAY68970"
FT CDS	1374..1823
FT	/*tag= b
FT	/note= "CpsF; encodes AAY68971"
FT CDS	1823..2317
FT	/*tag= c
FT	/note= "CpsG; encodes AAY68972"
FT CDS	3036..4202
FT	/*tag= d
FT	/note= "CpsH; encodes AAY68973; the nucleotides encoding
FT	amino acid 213 are not given"
FT CDS	4195..5163
FT	/*tag= e
FT	/note= "CpsI; encodes AAY68974"
FT CDS	5172..6143
FT	/*tag= f
FT	/transl_except= (pos: 5634..5636, aa: Xaa)
FT	/transl_except= (pos: 5637..5638, aa: Xaa)
FT	/note= "CpsJ; xaa is an unspecified amino acid; encodes
FT	AAY68975; nucleotides 5328-5330 encode an amino acid that
FT	is not included in the protein sequence"

FT	CDS		6156..6990	
FT		/tag=	"g"	
FT		/note=	"CpsIK; encodes AAY68976; no termination codon given"	
FT				
PN	MO200005378-A2.			
XX				
PF	19-JUL-1999;	99WO-NL000460.		
PR	22-JUL-1998;	98EP-00202465.		
DR	P-BSDR; AAY68970, AAY68971, AAY68972, AAY68973, AAY68974, AAY68975,			
XX	AAY68976.			
PA	(DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.			
PI	Smith HE;			
DR	WPI; 2000-1995I04/I7.			
XX	P-BSDR; AAY68970, AAY68971, AAY68972, AAY68973, AAY68974, AAY68975,			
XX	AAY68976.			
PT	New nucleic acid containing the capsular gene cluster of Streptococcus suis, used for serotype-specific detection and to generate antigens or mutants for vaccination.			
PS	Claim 5; Fig 4; 14pp; English.			
CC	The present sequence represents the capsular gene cluster of Streptococcus suis serotype 1. The genes in this cluster are involved in polysaccharide biosynthesis of capsular components and antigens. The CC proteins have glycosyltransferase activities (CpsII, CpsIII, CpsIV, CpsIX) and CP polymerase activities (CpsIH). The capsule confers bacterium resistance to complement-mediated opsonophagocytosis. The gene cluster is used as a source of probes and primers for serotype-specific detection of S. suis and is also useful for recombinant production of the proteins. The proteins are then useful for producing antigens that can be used in vaccines, for controlling or eradicating a Streptococcal disease, in humans or animals, e.g. against S. suis in pigs			
SQ	Sequence 6992 BP; 2513 A; 788 C; 1214 G; 2477 T; 0 U; 0 Other;			
	Query Match	100.0%; Score 6992; DB 3; Length 6992;		
	Best Local Similarity	100.0%; Pred. No. 0;		
	Matches 6992; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 ATCGGCCAAGCAAAATGGCATATTGATATATAGACAGTTGCATTTCTGCATCTTAA	60		
DB	1 ATCGGCCAAGCAAAATGGCATATTGATATATAGACAGTTGCATTTCTGCATCTTAA	60		
OY	61 CAAGTCATATACCAAATGCTGATTAAATCGTTTCGGAATTTTATCATATAGATGTTC	120		
DB	61 CAAGTCATATACCAAATGCTGATTAAATCGTTTCGGAATTTTATCATATAGATGTTC	120		
OY	121 ATTATTTTGCAATTTTATATCTCGTAGCCAGTTGAATTTGATATAGAGTAATCTGA	180		
DB	121 ATTATTTTGCAATTTTATATCTCGTAGCCAGTTGAATTTGATATAGAGTAATCTGA	180		
OY	181 TAGAGTTGAAAAAACATTACTATATAGTAATATATTCGCAATTTTCTTAGGAGATAT	240		
DB	181 TAGAGTTGAAAAAACATTACTATATAGTAATATATTCGCAATTTTCTTAGGAGATAT	240		
OY	241 CATTTTTGTTGGAGATAATTTGGCACTTTCAGAAGCTGTGCGCGTATTTCACTTAA	300		
DB	241 CATTTTTGTTGGAGATAATTTGGCACTTTCAGAAGCTGTGCGCGTATTTCACTTAA	300		
OY	301 TAAACTTCGTTTGTATACCTAATTTAAGCAATTTAAGCAATTAGCAATAGCTTTC	360		
DB	301 TAAACTTCGTTTGTATACCTAATTTAAGCAATTTAAGCAATTAGCAATAGCTTTC	360		
OY	361 TATTTTGCACATCTATCAAAAAAAGACGATTTTAATTCACACGCGTGAACGATGGGAAA	420		
DB	361 TATTTTGCACATCTATCAAAAAAAGACGATTTTAATTCACACGCGTGAACGATGGGAAA	420		

QY	421	ATATGCAAGTTTATTTGATACATTAACAATTCAAAAAATCTGTGCAATGGTG	480
DB	421	ATATGCAAGTTTATTTGATACATTAACAATTCAAAAAATCTGTGCAATGGTG	480
QY	481	TTTTGATACAGAAATAGATTAATAATTATCAATTACCGGCTATATTCGTGGAAG	540
DB	481	TTTTGATACAGAAATAGATTAATAATTATCAATTACCGGCTATATTCGTGGAAG	540
QY	541	AAGCTATGAGTTTCAACAAAGGAAGTGTGACACGCTCTTATTAATCTACAAGTG	600
DB	541	AAGCTATGAGTTTCAACAAAGGAAGTGTGACACGCTCTTATTAATCTACAAGTG	600
QY	601	AGTTTATGACGTAAAGCAATTCGTTTCAGATTTTGAGTTTGATATTGATGTAAGC	660
DB	601	AGTTTATGACGTAAAGCAATTCGTTTCAGATTTTGAGTTTGATATTGATGTAAGC	660
QY	661	TTGATATTATTCATTCGCTTTTACCTGCTGAAAAACAAAAAATCCAAAGCTGAGTG	720
DB	661	TTGATATTATTCATTCGCTTTTACCTGCTGAAAAACAAAAAATCCAAAGCTGAGTG	720
QY	721	ACCATAGCACTGTAACCTTTTCCACAATTTTATTAAGCCTAGTCATATCATGTAAC	780
DB	721	ACCATAGCACTGTAACCTTTTCCACAATTTTATTAAGCCTAGTCATATCATGTAAC	780
QY	781	GACTTTTGATATCTCGAGCGGTAGTCGGGTAAATTAATGTGTATATGTTCTATTT	840
DB	781	GACTTTTGATATCTCGAGCGGTAGTCGGGTAAATTAATGTGTATATGTTCTATTT	840
QY	841	TGTTAGTTCCATTAATTCGTAGAGATGTGAAACGGGCTATTTTGTCTAGAAACGAGTTG	900
DB	841	TGTTAGTTCCATTAATTCGTAGAGATGTGAAACGGGCTATTTTGTCTAGAAACGAGTTG	900
QY	901	GACAGATGAGCAATTAATTACATTCCTACAAAGTTTGATGATGATGTGATGCTGAGG	960
DB	901	GACAGATGAGCAATTAATTACATTCCTACAAAGTTTGATGATGATGTGATGCTGAGG	960
QY	961	ACGCGAAAAAGACTTGCTCAGCCAAAACCAATGCAAGGGTGGATATGTTTAAATGG	1020
DB	961	ACGCGAAAAAGACTTGCTCAGCCAAAACCAATGCAAGGGTGGATATGTTTAAATGG	1020
QY	1021	GAAAAACGATCCTAGAAATTAATCTCAATTGGAATTTCTATCGGAAAAACAGTTTGAACG	1080
DB	1021	GAAAAACGATCCTAGAAATTAATCTCAATTGGAATTTCTATCGGAAAAACAGTTTGAACG	1080
QY	1081	AGTTACCAAGTTTATAATGTTTAAATGCGCATATGAGTCTAGTTGGTACAGCTCAC	1140
DB	1081	AGTTACCAAGTTTATAATGTTTAAATGCGCATATGAGTCTAGTTGGTACAGCTCAC	1140
QY	1141	CTACAGTTGATGAAATTTGAAAAATATATCTCTGTGCAAAAGAGACGATGAGTTTAAAC	1200
DB	1141	CTACAGTTGATGAAATTTGAAAAATATATCTCTGTGCAAAAGAGACGATGAGTTTAAAC	1200
QY	1201	CAGGAAATACAGGCTCTGCGCAGGTTAGTGTCGTATGATTAATCACAGACTTGACACAGC	1260
DB	1201	CAGGAAATACAGGCTCTGCGCAGGTTAGTGTCGTATGATTAATCACAGACTTGACACAGC	1260
QY	1261	TAGTTGCGTTGACCTTAGCATATATGATTAATGGAAGCTATCTGTGAGATTTAAATTT	1320
DB	1261	TAGTTGCGTTGACCTTAGCATATATGATTAATGGAAGCTATCTGTGAGATTTAAATTT	1320
QY	1321	TATTTAAGACAGTGAAGTTGTATTTGTGAGAGAGGAAATGAATGAATATGAAG	1380
DB	1321	TATTTAAGACAGTGAAGTTGTATTTGTGAGAGAGGAAATGAATGAATATGAAG	1380
QY	1381	TTTGTGTGCTCGGTTCTTCAAGGGGACATTTGACTCACTTGATTTGTTAAACCGTTTT	1440
DB	1381	TTTGTGTGCTCGGTTCTTCAAGGGGACATTTGACTCACTTGATTTGTTAAACCGTTTT	1440
QY	1441	GGAAGGAGAAAGCGTTTTTGGGTAACTTTGATTAAGAGGATGCAAGAGCTTTTGA	1500
DB	1441	GGAAGGAGAAAGCGTTTTTGGGTAACTTTGATTAAGAGGATGCAAGAGCTTTTGA	1500

QY 1501 AGAATGAAAAATGATCATGTTACTTCCACAAATCGCAATCTCATTAATTTAGTGA 1560
DB 1501 AGAATGAAAAATGATCATGTTACTTCCACAAATCGCAATCTCATTAATTTAGTGA 1560
QY 1561 AAAATACTTTCTTAGCTTTCAAATTTTACGTGATGGAACCGAGTTATTTTCAT 1620
DB 1561 AAAATACTTTCTTAGCTTTCAAATTTTACGTGATGGAACCGAGTTATTTTCAT 1620
QY 1621 CTGTGCGCGCGTGTGTCCTCCCTCTTTTACATCGGAAAACCTATTTGAGCAAGACA 1680
DB 1621 CTGTGCGCGCGTGTGTCCTCCCTCTTTTACATCGGAAAACCTATTTGAGCAAGACA 1680
QY 1681 TTTATATGGAAGTATTTGATCGAGTTATTAATCTACATTTAATCTGAAAACTAGTTATC 1740
DB 1681 TTTATATGGAAGTATTTGATCGAGTTATTAATCTACATTTAATCTGAAAACTAGTTATC 1740
QY 1741 CCGTACAGATTTTATTTGTTCTGATGCGGAGAAATGGAAGGTAATCTTAATCTTA 1800
DB 1741 CCGTACAGATTTTATTTGTTCTGATGCGGAGAAATGGAAGGTAATCTTAATCTTA 1800
QY 1801 TTTAATCTGGGAGATTTTATTTATGATTTTGTACAGTAGGAACTCATGAAACAAGTT 1860
DB 1801 TTTAATCTGGGAGATTTTATTTATGATTTTGTACAGTAGGAACTCATGAAACAAGTT 1860
QY 1861 TTAATCGATGATAAAGAGATTGATTTATTTGAAAAAATGGAAGTATTAACCGACGAAT 1920
DB 1861 TTAATCGATGATAAAGAGATTGATTTATTTGAAAAAATGGAAGTATTAACCGACGAAT 1920
QY 1921 ATTTATTCAAACAGATATTTCTGATGATTTTCCAGAAATTTGCAAGTATTAATAATTTCT 1980
DB 1921 ATTTATTCAAACAGATATTTCTGATGATTTTCCAGAAATTTGCAAGTATTAATAATTTCT 1980
QY 1981 CAGTTTCAAAGAAATGGAACAATTTATTAACAATCGAAGTAGTTATTTGCCACGAGAG 2040
DB 1981 CAGTTTCAAAGAAATGGAACAATTTATTAACAATCGAAGTAGTTATTTGCCACGAGAG 2040
QY 2041 CCCCCGTACTTTATGATTCATTATTCGAAAGGAAAAAACAATATTTGTTTCTTAGACA 2100
DB 2041 CCCCCGTACTTTATGATTCATTATTCGAAAGGAAAAAACAATATTTGTTTCTTAGACA 2100
QY 2101 AAAAAAGTATGTAACATGTAATGATCATCAAGTAGTTTGTAGAAAGATTTTACA 2160
DB 2101 AAAAAAGTATGTAACATGTAATGATCATCAAGTAGTTTGTAGAAAGATTTTACA 2160
QY 2161 AGATTAATTAATTTTATTTATTAAGAAATATAGATGTTTGTGAATAATTTAGTGAAGT 2220
DB 2161 AGATTAATTAATTTTATTTATTAAGAAATATAGATGTTTGTGAATAATTTAGTGAAGT 2220
QY 2221 TTTCTAAGCAAACTAATCTTATCATCAATTAATTTTGTGAAAGATTTAAACAAT 2280
DB 2221 TTTCTAAGCAAACTAATCTTATCATCAATTAATTTTGTGAAAGATTTAAACAAT 2280
QY 2281 AGTTGAAAAATTTAATGAGATCAAGAAATGAAATATAAAAAAGATGATATTTGATAA 2340
DB 2281 AGTTGAAAAATTTAATGAGATCAAGAAATGAAATATAAAAAAGATGATATTTGATAA 2340
QY 2341 TGGCTATCATTAATTTTCTCAAGTTTACTGAGAGGATACAGATATTTATCATCTTCT 2400
DB 2341 TGGCTATCATTAATTTTCTCAAGTTTACTGAGAGGATACAGATATTTATCATCTTCT 2400
QY 2401 CTGAGAGATGAGACCATAGTCTTCTGAGAAATCTGTATTAATTTTAAATATTC 2460
DB 2401 CTGAGAGATGAGACCATAGTCTTCTGAGAAATCTGTATTAATTTTAAATATTC 2460
QY 2461 TCAAGATTTATGTTGAATTTTCAAAAAGATGAGCAAAAATATTAAGAAATAGATATA 2520
DB 2461 TCAAGATTTATGTTGAATTTTCAAAAAGATGAGCAAAAATATTAAGAAATAGATATA 2520
QY 2521 TGAACGAGTTAAATGTTACGATTTATTTCTTAATATATCAAAAAAATCTATGATATGT 2580
DB 2521 TGAACGAGTTAAATGTTACGATTTATTTCTTAATATATCAAAAAAATCTATGATATGT 2580
QY 2581 ACTGTTAGAAATTTTATTAAGATGATCGAGCTTTGAAATACATTTACAAGATTTGTT 2640

DB 2581 ACTGTTAGAAATTTTATTAAGAAATGATCGAGCTTTGAAATACATTTACAAGATTTGTT 2640
QY 2641 GTTATTTAGTAAGATTAATAAACAATGCTTAAGATATAGATTTGGTCTTAATTTGGCTTCG 2700
DB 2641 GTTATTTAGTAAGATTAATAAACAATGCTTAAGATATAGATTTGGTCTTAATTTGGCTTCG 2700
QY 2701 CTTCCACATGATTTTGTGCAATCTTTTATCAAAATGAAACGAAACAGCTTATTTATTT 2760
DB 2701 CTTCCACATGATTTTGTGCAATCTTTTATCAAAATGAAACGAAACAGCTTATTTATTT 2760
QY 2761 AAGTAATCTAAATGTCAGATGAACTATTTTATACAGCAATTAATGAAAAATATGAATTT 2820
DB 2761 AAGTAATCTAAATGTCAGATGAACTATTTATACAGCAATTAATGAAAAATATGAATTT 2820
QY 2821 TCAAAATGATTTATCTAAATATGGAATTTAAAGATATTAAGTGAAGAAAAATCAACTCT 2880
DB 2821 TCAAAATGATTTATCTAAATATGGAATTTAAAGATATTAAGTGAAGAAAAATCAACTCT 2880
QY 2881 TCTCTATTTGCTTTACAGATGATCTATGATGAAATTTGCTAAATGCAAGAAATTTAGT 2940
DB 2881 TCTCTATTTGCTTTACAGATGATCTATGATGAAATTTGCTAAATGCAAGAAATTTAGT 2940
QY 2941 TTTTATTTGCTAGAAAGTTAAAAATGAAATPAATCTAAATTTTAAAGAAATTTATTA 3000
DB 2941 TTTTATTTGCTAGAAAGTTAAAAATGAAATPAATCTAAATTTTAAAGAAATTTATTA 3000
QY 3001 AAAAAATAAATAGTTGATTTTGTGAGAGTAATGATTTTAAATTAATTAATATAC 3060
DB 3001 AAAAAATAAATAGTTGATTTTGTGAGAGTAATGATTTTAAATTAATTAATTAATATAC 3060
QY 3061 CGGAATATTTTATTTTATGATCTCGGTGATTTATTTTATTCGAGCAAAAGTATG 3120
DB 3061 CGGAATATTTTATTTTATGATCTCGGTGATTTATTTTATTCGAGCAAAAGTATG 3120
QY 3121 TATTTTATTAATTTTATGAAATTTATTTATTTTATTCATATPAATTTTGAACCTAAGC 3180
DB 3121 TATTTTATTAATTTTATGAAATTTATTTATTTTATTCATATPAATTTTGAACCTAAGC 3180
QY 3181 TAAATTAATAAATAAGAAATTTATTTGTTTATTAATGCTATATTAATGTTTGTTCAG 3240
DB 3181 TAAATTAATAAATAAGAAATTTATTTGTTTATTAATGCTATATTAATGTTTGTTCAG 3240
QY 3241 TAGTCAAGATAGTTTGTGAAATTAATTTTGAAGATTAATGCAATTTTACTGCTC 3300
DB 3241 TAGTCAAGATAGTTTGTGAAATTAATTTTGAAGATTAATGCAATTTTACTGCTC 3300
QY 3301 CCATTAATTTGATTTATGCAATATGATATTAATTTGATTCATTTATTAATTTGATTT 3360
DB 3301 CCATTAATTTGATTTATGCAATATGATATTAATTTGATTCATTTATTAATTTGATTT 3360
QY 3361 ATAAAAATTAATAAATAAGATTTATTTGTTTATTAATGCTATATTTGATTT 3420
DB 3361 ATAAAAATTAATAAATAAGATTTATTTGTTTATTAATGCTATATTTGATTT 3420
QY 3421 TGTATTAATTAATTAATGGAAGATTTGATTTTATGACACACTTATAGAGCTAG 3480
DB 3421 TGTATTAATTAATTAATGGAAGATTTGATTTTATGACACACTTATAGAGCTAG 3480
QY 3481 ACTATCTTAATTAAGGCGCTCAAAAAGGTTGCTGCTTATGAACTATCTAGCTTAA 3540
DB 3481 ACTATCTTAATTAAGGCGCTCAAAAAGGTTGCTTATGAACTATCTAGCTTAA 3540
QY 3541 ATACCACTAATTAATTAATGCTTAAATCGTTAATCTTTGACCTTATTAATAAATATGC 3600
DB 3541 ATACCACTAATTAATTAATGCTTAAATCGTTAATCTTTGACCTTATTAATAAATATGC 3600
QY 3601 AACAAATTTTCTGCTGCTGCTTATTAACGATCTATTAAGTGAATGAGATTTG 3660
DB 3601 AACAAATTTTCTGCTGCTGCTTATTAACGATCTATTAAGTGAATGAGATTTG 3660
QY 3661 GTAGTTATGCTAGCAATTAATTAATTAATGCTTGTATGAGATATTAAGTGAAT 3720

Db 3661 GTAGTTTATCGTAGCAATTAATTATATGCTGTGTATGAGATATATAGTGAGAAAT 3720
QY 3721 TTGCTGTGATAAAAAGCTATATAGTATATTTGTAATACTATTTATTTAAATACAG 3780
Db 3721 TTGCTGTGATAAAAAGCTATATAGTATATTTGTAATACTATTTATTTAAATACAG 3780
QY 3781 AATTCGTTTACCAAGAAATTTTGGCTGTTTAAATCTAGAGAAATCAAGTAAAGAGCTA 3840
Db 3781 AATTCGTTTACCAAGAAATTTTGGCTGTTTAAATCTAGAGAAATCAAGTAAAGAGCTA 3840
QY 3841 GATTTATTTATTTTCAAGAGAGTATGATTAAGTATTTAGAAAACAATATTTTATTTGGAT 3900
Db 3841 GATTTATTTATTTTCAAGAGAGTATGATTAAGTATTTAGAAAACAATATTTTATTTGGAT 3900
QY 3901 ATGGAATATCCGAAATATTCAGTTACGGAGACTTGGCTCGAGAGTCATTCAGGCTATATAT 3960
Db 3901 ATGGAATATCCGAAATATTCAGTTACGGAGACTTGGCTCGAGAGTCATTCAGGCTATATAT 3960
QY 3961 CATTTTATTTAATAATGAGAAATAGTTGGGTGATTTTACTGATGTTTTCTTTTTTTATG 4020
Db 3961 CATTTTATTTAATAATGAGAAATAGTTGGGTGATTTTACTGATGTTTTCTTTTTTTATG 4020
QY 4021 TTTTAAAAAAAAGTATGAGTTAAATGCGGAAAAAGACATTTTATTTATTTACATCAATAG 4080
Db 4021 TTTTAAAAAAAAGTATGAGTTAAATGCGGAAAAAGACATTTTATTTATTTACATCAATAG 4080
QY 4081 CCATATTTTTCATATATGAACAATAGATCCGATTTATTTATTTATTTAGTACTATTTCTTT 4140
Db 4081 CCATATTTTTCATATATGAACAATAGATCCGATTTATTTATTTATTTAGTACTATTTCTTT 4140
QY 4141 CTTCAATAGGTATTTGGAAATATATATTAATTTTAAAAAGATATGAGACAAAAATGAAAT 4200
Db 4141 CTTCAATAGGTATTTGGAAATATATATTAATTTTAAAAAGATATGAGACAAAAATGAAAT 4200
QY 4201 GATTTAATTTCAGTTATTTGACCAATTTATATATGTCGAAATTTCTTGATTAATGATAT 4260
Db 4201 GATTTAATTTCAGTTATTTGACCAATTTATATATGTCGAAATTTCTTGATTAATGATAT 4260
QY 4261 AACAGATTTATTAACCAAAACATATCTAATTTTAGAGGTTATCTCGTAAATGATGAGAT 4320
Db 4261 AACAGATTTATTAACCAAAACATATCTAATTTTAGAGGTTATCTCGTAAATGATGAGAT 4320
QY 4321 ACTGATGATTTCTGAGAAATTTTGGCTTAACTATATGAGAAACGATGAGAAATTTAAATAT 4380
Db 4321 ACTGATGATTTCTGAGAAATTTTGGCTTAACTATATGAGAAACGATGAGAAATTTAAATAT 4380
QY 4381 TACAGAAAAATTAATGCGCGCTTAGCAGATGCTCGAAATTTTGGACTTAAAGATGACACA 4440
Db 4381 TACAGAAAAATTAATGCGCGCTTAGCAGATGCTCGAAATTTTGGACTTAAAGATGACACA 4440
QY 4441 GGTAAATATATGCTTTGTGCGATTTGATGACTATATAGAAATGATGCAATGTTGAGAGA 4500
Db 4441 GGTAAATATATGCTTTGTGCGATTTGATGACTATATAGAAATGATGCAATGTTGAGAGA 4500
QY 4501 ATGATGATATATATATGAGTATATGCGATATAGCAGATAGAGATTTTGTGTTAGTA 4560
Db 4501 ATGATGATATATATATGAGTATATGCGATATAGCAGATAGAGATTTTGTGTTAGTA 4560
QY 4561 GACGAAAAACGGGTATCAAAAGAAAAAAGAAATGTAATTTTCAATGCTTAAACAGAGAA 4620
Db 4561 GACGAAAAACGGGTATCAAAAGAAAAAAGAAATGTAATTTTCAATGCTTAAACAGAGAA 4620
QY 4621 GAGACTGTAAAGAAATTTTGTGCGATCTAATATAGAAAAATATGTTGGTGCAGCTT 4680
Db 4621 GAGACTGTAAAGAAATTTTGTGCGATCTAATATAGAAAAATATGTTGGTGCAGCTT 4680
QY 4681 TATTCAGAGATATATTAAGATATTAATATTTCCAAATTTAATTAAGAGATTTGTGAG 4740
Db 4681 TATTCAGAGATATATTAAGATATTAATATTTCCAAATTTAATTAAGAGATTTGTGAG 4740
QY 4741 GATTTGCTTTTAAATTTGAGAGCTTGTGAACATGTAAACGTGATGATGATCTAGA 4800
Db 4741 GATTTGCTTTTAAATTTGAGAGCTTGTGAACATGTAAACGTGATGATGATCTAGA 4800

QY 4801 GAATATTAATTAATTAATGATGATTCGTAAACAGTTGCTTATTAATCAGAAATTTCTTATA 4860
Db 4801 GAATATTAATTAATTAATGATGATTCGTAAACAGTTGCTTATTAATCAGAAATTTCTTATA 4860
QY 4861 AATTAATTAATGATTTAGTCACAAGATTTGAGAAATTTACCCCTTAAAGTTAAAAAGAGCTT 4920
Db 4861 AATTAATTAATGATTTAGTCACAAGATTTGAGAAATTTACCCCTTAAAGTTAAAAAGAGCTT 4920
QY 4921 AGTCATTTATTTGATGACAAAGTTATTAAGAGAGTTAAATGTTTAAACAAATGATAT 4980
Db 4921 AGTCATTTATTTGATGACAAAGTTATTAAGAGAGTTAAATGTTTAAACAAATGATAT 4980
QY 4981 TCAACAGATTTGTTGATATAGATTTCTTCCAAATTTAGAGCTTTATCGAAAAAGAAAT 5040
Db 4981 TCAACAGATTTGTTGATATAGATTTCTTCCAAATTTAGAGCTTTATCGAAAAAGAAAT 5040
QY 5041 CGTATATATCCATTTATTTAAACGAAAAAATTTTATCAAGAAAGATTTAGTTAGCTTG 5100
Db 5041 CGTATATATCCATTTATTTAAACGAAAAAATTTTATCAAGAAAGATTTAGTTAGCTTG 5100
QY 5101 TATTTGATGAAATTTTCCGCTAAACATATATGTAATGTTATATTAAGAAATTTCAAAAGCAG 5160
Db 5101 TATTTGATGAAATTTTCCGCTAAACATATATGTAATGTTATATTAAGAAATTTCAAAAGCAG 5160
QY 5161 TAGAGGTAAAAATGATAAAAATTTAGTTATTTGTTCCAGTTTATTAATGATAAATATTT 5220
Db 5161 TAGAGGTAAAAATGATAAAAATTTAGTTATTTGTTCCAGTTTATTAATGATAAATATTT 5220
QY 5221 TAAGTATGTTATAGAAACATTTATTAATCAAAAATTAATAAAAATATAGAAATTTATTTGA 5280
Db 5221 TAAGTATGTTATAGAAACATTTATTAATCAAAAATTAATAAAAATATAGAAATTTATTTGA 5280
QY 5281 TAGATATGCTCTGTAGATGATTTCTGTAATATATGCAAGAAATATGCAAGAAAGATAT 5340
Db 5281 TAGATATGCTCTGTAGATGATTTCTGTAATATATGCAAGAAATATGCAAGAAAGATAT 5340
QY 5341 AAAAGGTAAAAATTTTTCATTAATCATATGAGATATCAAAATGCTAGAAATCATAGAA 5400
Db 5341 AAAAGGTAAAAATTTTTCATTAATCATATGAGATATCAAAATGCTAGAAATCATAGAA 5400
QY 5401 TAAAGCGGAGTACAGCTGAATATATATGTTTGTGACCTGATGATGATGTTGATAGTA 5460
Db 5401 TAAAGCGGAGTACAGCTGAATATATATGTTTGTGACCTGATGATGATGTTGATAGTA 5460
QY 5461 GATTAAGTAAAAATTTATATTTTAAATATTAAGAGTAAATGATTTATCTGTTGTT 5520
Db 5461 GATTAAGTAAAAATTTATATTTTAAATATTTTAAAGTAAATGATTTATCTGTTGTT 5520
QY 5521 TGTACGCTACTTTTTCAGAAAAATATTAATATTTTGAAGTAAATCCAAATATTTGAT 5580
Db 5521 TGTACGCTACTTTTTCAGAAAAATATTAATATTTTGAAGTAAATCCAAATATTTGAT 5580
QY 5581 TTGAAGCAATTAATACCGTCAGACATAGGAGAAAAAATTTTATGAATTTGATATTA 5640
Db 5581 TTGAAGCAATTAATACCGTCAGACATAGGAGAAAAAATTTTATGAATTTGATATTA 5640
QY 5641 ATATATATTTTTCAGCTCTGTTGTAACTATATTAAGAAAGATACATACAGATCTTT 5700
Db 5641 ATATATATTTTTCAGCTCTGTTGTAACTATATTAAGAAAGATACATACAGATCTTT 5700
QY 5701 TTCAAGAGATCAATGTTTGAAGAAAGATTTACTTTTATATCTGATTTATTTAAAGATA 5760
Db 5701 TTCAAGAGATCAATGTTTGAAGAAAGATTTACTTTTATATCTGATTTATTTAAAGATA 5760
QY 5761 TAGATGAGATTTAGTTATTTGACCTGAACATCTTTATTTTATAGAGAGGTATCTAAAGTA 5820
Db 5761 TAGATGAGATTTAGTTATTTGACCTGAACATCTTTATTTTATAGAGAGGTATCTAAAGTA 5820
QY 5821 CAGTAAATTTCTTTTAAAGAGAGTGTGTTTTCGAAATGAAAAATTTGCAAAAACAAGTGA 5880
Db 5821 CAGTAAATTTCTTTTAAAGAGAGTGTGTTTTCGAAATGAAAAATTTGCAAAAACAAGTGA 5880

QY 5881 TAGATATGTTTAAGCAAAATATATGTCGAGATTTTGAAGTACATATGTTAAAGATACTA 5940
 DB 5881 TAGATATGTTTAAGCAAAATATATGTCGAGATTTTGAAGTACATATGTTAAAGATACTA 5940
 QY 5941 TACGTTGGCAAGTATTTTATTTATGCTTACTAATGTTTAAATACGAAAAAGCTATTTT 6000
 DB 5941 TACGTTGGCAAGTATTTTATTTATGCTTACTAATGTTTAAATACGAAAAAGCTATTTT 6000
 QY 6001 TTGACAAATTTTATTTTATTTAGAAATCTTATATAAAAAATTTTATTTTAACTGTTAAAG 6060
 DB 6001 TTGACAAATTTTATTTTATTTAGAAATCTTATATAAAAAATTTTATTTTAACTGTTAAAG 6060
 QY 6061 TATCTAACAAATTTCTTGTCTTAAATTTTGTATAGAAATTTGTTTCGACAAAGTTT 6120
 DB 6061 TATCTAACAAATTTCTTGTCTTAAATTTTGTATAGAAATTTGTTTCGACAAAGTTT 6120
 QY 6121 TTTAAAAATATATATGTTTAAATATGGAAGATATCATGATACTATTAGTAAATTTTCTA 6180
 DB 6121 TTTAAAAATATATATGTTTAAATATGGAAGATATCATGATACTATTAGTAAATTTTCTA 6180
 QY 6181 TAAATGTACTAT 6240
 DB 6181 TAAATGTACTAT 6240
 QY 6241 ATCAGACCTTAAACATATATAGATTTCTTGTGGAATGACGGATAGCGATTAATTCGG 6300
 DB 6241 ATCAGACCTTAAACATATATAGATTTCTTGTGGAATGACGGATAGCGATTAATTCGG 6300
 QY 6301 AAGAAATTTGTTTATGATATGCGAAGAAATATGTCGATTCGTTATTTTAAAAAGAGA 6360
 DB 6301 AAGAAATTTGTTTATGATATGCGAAGAAATATGTCGATTCGTTATTTTAAAAAGAGA 6360
 QY 6361 ACGCGCGGCTATCATGATGCCCTTATATATGTCGATTAATGCGGCAAGGATCTACTTAG 6420
 DB 6361 ACGCGCGGCTATCATGATGCCCTTATATATGTCGATTAATGCGGCAAGGATCTACTTAG 6420
 QY 6421 CTTTATATGACTAGATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 6480
 DB 6421 CTTTATATGACTAGATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 6480
 QY 6481 TTGAGAGAGAAATGCGCTTGTGCGAGTTCGTTATGATAGGATGATGCTTCGAGGC 6540
 DB 6481 TTGAGAGAGAAATGCGCTTGTGCGAGTTCGTTATGATAGGATGATGCTTCGAGGC 6540
 QY 6541 ATTTCTTAAACAGAGAGCGCTTCTTCAAAATCAGGCTGTTCTGAGCGGAGGATGTTT 6600
 DB 6541 ATTTCTTAAACAGAGAGCGCTTCTTCAAAATCAGGCTGTTCTGAGCGGAGGATGTTT 6600
 QY 6601 GTAAAAAGCTGTAAGAGCGGAGTGTGATGCTTGTGAGCGCTGTAATTAATCTATA 6660
 DB 6601 GTAAAAAGCTGTAAGAGCGGAGTGTGATGCTTGTGAGCGCTGTAATTAATCTATA 6660
 QY 6661 AAAAAAGACTATTTGAAGATTTTGAATTTGAAAAAGGTAAGATTCATGAAGATGAATACT 6720
 DB 6661 AAAAAAGACTATTTGAAGATTTTGAATTTGAAAAAGGTAAGATTCATGAAGATGAATACT 6720
 QY 6721 TCACTTATGCTTGTCTTATGAGTTAGAAAAAGTTCATATAGTTAAGAGTGTCTGTACT 6780
 DB 6721 TCACTTATGCTTGTCTTATGAGTTAGAAAAAGTTCATATAGTTAAGAGTGTCTGTACT 6780
 QY 6781 AATATGTGACCGAAGAAATATATATCAACTTCTACATGATGACCATGCTTCAT 6840
 DB 6781 AATATGTGACCGAAGAAATATATATCAACTTCTACATGATGACCATGCTTCAT 6840
 QY 6841 GCGTACTGAAATTTCAAAATGAAGATGACCTTCTATGAAAGTATGAGAGATTAAGAGC 6900
 DB 6841 GCGTACTGAAATTTCAAAATGAAGATGACCTTCTATGAAAGTATGAGAGATTAAGAGC 6900
 QY 6901 TCTTACTAGAGTGTATGCTTCATTTTATGCTTGTCTGTTTGTTTTATGAGCAAAATATA 6960
 DB 6901 TCTTACTAGAGTGTATGCTTCATTTTATGCTTGTCTGTTTGTTTTATGAGCAAAATATA 6960
 QY 6961 ATCATTGTTGAGCAAAACAGCAAAAGAGCTT 6992

DB 6961 ATCATTGTTGAGCAAAACAGCAAAAGAGCTT 6992
 RESULT 2
 AAZ60929
 ID AAZ60929 standard; DNA; 26281 BP.
 XX
 AC AAZ60929;
 DT 30-MAY-2000 (first entry)
 XX
 DE Nucleotide sequence of a capsular gene cluster of *S. suis* serotype 2.
 XX
 KW Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
 KW capsular component; antigen; regulation; chain length determination;
 KW complement-mediated opsonophagocytosis; serotype-specific detection;
 KW antigen; vaccine; Streptococcal diseases; ORF 2Z; ORF 2Y; ORF 2Z; Cps2A;
 KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;
 KW Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T; ss.
 OS
 XX Streptococcus suis.
 FH Key
 FT CDS 2..721
 FT Location/Qualifiers
 FT /*tag= a
 FT /note= "ORF 2Z; encodes AAY68950"
 FT complement (822..2079)
 FT /*tag= b
 FT /note= "ORF 2Y; encodes AAY68951"
 FT 2202..2936
 FT /*tag= c
 FT /tranel except= (pos: 2916..2918, aa: Xaa)
 FT /note= "ORF 2X; Xaa is an unspecified amino acid; encodes
 FT AAY68952"
 FT 3041..4486
 FT /*tag= d
 FT /note= "Cps2A; encodes AAY68953"
 FT 4504..5193
 FT /*tag= e
 FT /note= "Cps2B; encodes AAY68954"
 FT 5203..5880
 FT /*tag= f
 FT /note= "Cps2C; encodes AAY68955"
 FT 5919..6650
 FT /*tag= g
 FT /note= "Cps2D; encodes AAY68956"
 FT 6675..8054
 FT /*tag= h
 FT /note= "Cps2E; encodes AAY68957"
 FT 8089..9258
 FT /*tag= i
 FT /note= "Cps2F; encodes AAY68958"
 FT 9262..10419
 FT /*tag= j
 FT /tranel except= (pos: 10057..10059, aa: Xaa)
 FT /note= "Cps2G; Xaa is an unspecified amino acid; encodes
 FT AAY68959"
 FT 10808..12176
 FT /*tag= k
 FT /tranel except= (pos: 11963..11965, aa: Xaa)
 FT /note= "Cps2H; Xaa is an unspecified amino acid; encodes
 FT AAY68960"
 FT 12213..13445
 FT /*tag= l
 FT /note= "Cps2I; encodes AAY68961"
 FT 13583..14581
 FT /*tag= m
 FT /note= "Cps2J; encodes AAY68962"
 FT 14574..15578
 FT /*tag= n
 FT /note= "Cps2K; encodes AAY68963"
 FT 18401..18904
 FT CDS

```
FT      /tag= 0
FT      /transl_except= (pos: 18755..18957, aa: Ile)
FT      /transl_except= (pos: 18770..18772, aa: Leu)
FT      /transl_except= (pos: 18778..18778, aa: Ile)
FT      /transl_except= (pos: 18836..18838, aa: Pro)
FT      /transl_except= (pos: 18880..18892, aa: Thr)
FT      /transl_except= (pos: 18896..18936, aa: Pro)
FT      /transl_except= (pos: 18937..18937, aa: Pro)
FT      /transl_except= (pos: 18953..18955, aa: Leu)
FT      /transl_except= (pos: 18968..18970, aa: Leu)
FT      /transl_except= (pos: 18971..18973, aa: Trp)
FT      /transl_except= (pos: 18977..18979, aa: Ser)
FT      /transl_except= (pos: 18980..18982, aa: Gln)
FT      /transl_except= (pos: 18989..18991, aa: Asn)
FT      /note= "Cps20; Xaa is an unspecified amino acid; encodes
FT      AAY68964"
FT      CDS
FT      20327..21343
FT      /tag= p
FT      /note= "Cps2P; encodes AAY68965"
FT      21355..21867
FT      /tag= q
FT      /note= "Cps2Q; encodes AAY68966"
FT      21933..22487
FT      CDS
FT      /tag= r
FT      /note= "Cps2R; encodes AAY68967"
FT      22501..23127
FT      /tag= s
FT      /note= "Cps2S; encodes AAY68968"
FT      23136..24368
FT      /tag= t
FT      /note= "Cps2T; encodes AAY68969"
FT      MO200005378-A2.
XX      03-FEB-2000.
XX      19-JUL-1999; 99WO-NL000460.
XX      22-JUL-1998; 98EP-00202465.
XX      22-JUL-1998; 98EP-00202467.
XX      (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX      Smith HE;
XX      MPI; 2000-195104/17.
XX      P-PSDB; AAY68950, AAY68951, AAY68952, AAY68953, AAY68954, AAY68955,
XX      AAY68956, AAY68957, AAY68958, AAY68959, AAY68960, AAY68961, AAY68962,
XX      AAY68963, AAY68964, AAY68965, AAY68966, AAY68967, AAY68968, AAY68969.
XX      New nucleic acid containing the capsular gene cluster of Streptococcus
XX      suis, used for serotype-specific detection and to generate antigens or
XX      mutants for vaccination.
XX      Claim 4; Fig 3; 144p; English.
XX      The present sequence represents the capsular gene cluster of
XX      Streptococcus suis serotype 2. The genes in this cluster are involved in
XX      polysaccharide biosynthesis of capsular components and antigens. The
XX      proteins are involved in regulation (CpsA), chain length determination
XX      (CpsB, CpsC), export (CpsE), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
XX      CpsI, CpsK). The capsule confers bacterium resistance to complement-
XX      mediated opsonophagocytosis. The gene cluster is used as a source of
XX      probes and primers for serotype-specific detection of S. suis and is also
XX      useful for recombinant production of the proteins. The proteins are then
XX      CC useful for producing antigens that can be used in vaccines, for
XX      CC controlling or eradicating a Streptococcal disease, in humans or animals,
XX      CC e.g. against S. suis in pigs
XX      Sequence 26281 BP; 8252 A; 4158 C; 5488 G; 8383 T; 0 U; 0 Other;
XX      Query Match 17.1%; Score 1195.8; DB 3; Length 26281;
XX      Best Local Similarity 78.9%; Pred. No. 1.3e-151;
```

```
Matches 1437; Conservative 0; Mismatches 382; Indels 2; Gaps 1;
QY      5172 ATGATGAAATAGTGTATGTTCCAGTTATTAATGATGATTAATTTAAAGTGTGT 5231
DB      13583 ATGGAAGAAAGTCAGATTAATGTTACCTATTTTAAATACGGAAGACTAAAGAGTGT 13642
QY      5232 ATGGAAGATTTATTAACAAATTTAATAAATATGAAATTTATGATGATGAGCG 5291
DB      13643 TTAGATGATTTATTTCCCAATCGTATCAATCTGAGATTTCTTGTATGATGACGGT 13702
QY      5292 TCTGTAGATGATTTGCTTAATATGCAAGAAATGCAAGAAAGATTTAAAGATTA 5351
DB      13703 TCTTGATGATTTCAACGGAATATGTTGGAATACGACGACGATGATGATTA 13762
QY      5352 ATTTTTCATCAATCATAGTGAAGATCAAAATGCTAGAAATCATGAAATTAAGCGGAGT 5411
DB      13763 CTTTCCGGTTACCAATATGTTGGTTCCTCAACGCAAGGAATTTACGGATTCAAAATATAC 13822
QY      5412 ACAGCTGAATATTTATGTTTGTGACTGATGATGTTGTTGATGATGATGATGATGAA 5471
DB      13823 ACAGCAATTTATTTATTTGTTGATGATGATGATGATGATGATGATGATGATGATG 13882
QY      5472 AATTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5531
DB      13883 TCTTTATACCTGTTTAAAGAAATGATGATGATTTGTCGGAGAGTTACTGCTACT 13942
QY      5532 TTTTCAGAAATATTAATTAATTTTGAAGTAAATCAAAATTTGATTTTGAAGCAAT 5591
DB      13943 TTTGATGAAATTTATCAAGAAATCTGAGCTGCAAAAGTCAAAATTTGATTTGGAAGATA 14002
QY      5592 AATACCGTCAGACATCGGAGAAAAATTTTATGAAATTTGATGATTTATTAATTTT 5651
DB      14003 AAAGAGTCGAGACTTGAAGAAATGAAATTTTCCCAATCATTAATGAGCGGTATCTTT 14062
QY      5652 TCTACTCTGTTTGTAACTATTAAGAAAGATACATACAGATCTTTTCAAGAAAT 5711
DB      14063 AATAGCCCTTTGTCGAACTTTATTAAGAAATATTAATTAACCAAGCTTTTGACATGAA 14122
QY      5712 CAATGTTAGAGAGAGATTTACTTTTAAATGCAATTTTAAAGAAATATGATGATGAT 5771
DB      14123 CAGTGTATTGAGAGAGACTTATTTATTTAAATTTTAAAGAAATTTAAAGAAATTTAAAG 14182
QY      5772 AGTTATTTGACGACATCTTTATTTTATGAGAGGATTTCTAATGATCAATTAATTC 5831
DB      14183 CGCTATGTTAAGAAATCTTTATTTTGCAGAAAGATTTCAAAATCATTAATGAGCGGTATCTTT 14242
QY      5832 TTTAAAGAGGAGTGTGTTTTCGAATGGAATTTGCAAAACCAAGTATGATTTGTT 5891
DB      14243 TTTAAATATGATGTTTATTTCAATTAAGAAATTTGAAAGAAATTTGATTTGTTT 14302
QY      5892 AAGCAATATATGCTGAGATTTTTCGATCAATGTTTAAAGATTTAATGCTGAGCA 5951
DB      14303 GTTAAATATTTGTTGAGCAATATGATTTTCTGTTTAAAGAGCGTACAGTGCGAT 14362
QY      5952 GATTTTATTTATGCTTACTAATGTTTAAATACGAAACAGTCATTTTGACAAATTT 6011
DB      14363 ATTAATTTATTTATGCTTATTTAATGTTCAAAATGGAAGAAATGCTTCCAAAGAAATG 14422
QY      6012 TTAATTTTAAAGATCTTTAATAAATATTTATTTAATCTGTTAAAGATATCAAA 6071
DB      14423 CATATATTTT--AAGTATTTATCAAAATGCGATCTTTAGTACTCTAAGATTTAAACGA 14480
QY      6072 AATTCCTTGTCAAAATTTTGTATTAAGAAATGTTTCGAACCAAGTTTAAAGAAAT 6131
DB      14481 CGTCTCTGTTTAAAGAAATATGTAATTAATTTGTTGTAATTTGTTTAAATTT 14540
QY      6132 TTAATGTTATTAAGAGATATCATGATATCTATTAATTTCTATTAATTTGATCT 6191
DB      14541 TTTTAATATCTTTAATTTAGGAAGAAATATGATTTAATCTATCATCTGTCCTCA 14600
QY      6192 ATATATTAATGTAAGAAATTTTATCTAAATGATATGATGATGATGATGATGATGATGATG 6251
DB      14601 ATTTACAAATGTTGAACAAATATCTATCAAGTATTAATGATGATGATGATGATGATGATGATG 14660
```

OY	6252	AAACCTATAGAGATTCCTTCGCGTAATGACGGTAGTACGGATATTCGGAAGAAATTTGT	6311
Db	14661	AAACCTATAGAGATTCCTTCGCGTAATGACGGTAGTACGGATATTCGGAAGAAATTTGT	14720
OY	6312	TTAGCATATGCGAAGAAAGATAGTCGGATTCGTTATTTTAAAAAGAGAACGGCGGCTA	6371
Db	14721	TTAGCATATGCGAAGAAAGATAGTCGGATTCGTTATTTTAAAAAGAGAACGGCGGCTA	14780
OY	6372	TCAGATGCCCTAATTTATGGCATAGTCGGCGCAAGGTGACATCTTAGCTTTTATAGAC	6431
Db	14781	TCAGATGCCCTAATTTATGGCATAGTCGGCGCAAGGTGACATCTTAGCTTTTATAGAC	14840
OY	6432	TCAGATGATTTTATTCATTCGGAGTTATCCAAAGTTTACACGAACCAATTGAGAGAG	6491
Db	14841	TCAGATGATTTTATTCATTCGGAGTTATCCAAAGTTTACACGAACCAATTGAGAGAG	14900
OY	6492	AATGCCCTTGGGAGTTGCTGTTATGATAGGGTAGATGCTTCGGGGCATTTCTTACAC	6551
Db	14901	AATGCCCTTGGGAGTTGCTGTTATGATAGGGTAGATGCTTCGGGGCATTTCTTACAC	14960
OY	6552	GCAAGCCGCTTCTCAAAATCAGGCTGTTCTGAGCGGACGGAATGTTGTAAAAAGCTG	6611
Db	14961	GCAAGCCGCTTCTCAAAATCAGGCTGTTCTGAGCGGACGGAATGTTGTAAAAAGCTG	15020
OY	6612	CTAAGAGGCGATGTCATCGCTTTGGTGGGCGCTGTATTAATCTATATAAAAAGACAT	6671
Db	15021	CTAAGAGGCGATGTCATCGCTTTGGTGGGCGCTGTATTAATCTATATAAAAAGACAT	15080
OY	6672	TTTGAAGATTTTCGATTTGAAAAAGGTAGATTCATGAGATGAATATCTTCACTTATCGC	6731
Db	15081	TTTGAAGATTTTCGATTTGAAAAAGGTAGATTCATGAGATGAATATCTTCACTTATCGC	15140
OY	6732	TTGCTCTATGAGTTAGAAAAAGTTGCATATGTTAAGAGTGCTTGTACTATATGTTGAC	6791
Db	15141	TTGCTCTATGAGTTAGAAAAAGTTGCATATGTTAAGAGTGCTTGTACTATATGTTGAC	15200
OY	6792	CGAGAAATAGTATCACAACTTTCAGAGATGACATCGCTTCATTTGCCATCTGGAA	6851
Db	15201	CGAGAAATAGTATCACAACTTTCAGAGATGACATCGCTTCATTTGCCATCTGGAA	15260
OY	6852	TTTCAAAATGAAAGCAATGGAATCTCTATGAAAGTAGAGAGATTAAGAGCTTTACTAGAG	6911
Db	15261	TTTCAAAATGAAAGCAATGGAATCTCTATGAAAGTAGAGAGATTAAGAGCTTTACTAGAG	15320
OY	6912	TGTTATGCTTCATTTTATAGCCTTTGCTGTTTGTTTTATAGGCAATATATCATTTGCTTG	6971
Db	15321	TGTTATGCTTCATTTTATAGCCTTTGCTGTTTGTTTTATAGGCAATATATCATTTGCTTG	15380
OY	6972	AGCAAAACGCAAAAAGAGCTT 6992	
Db	15381	AGCAAAACGCAAAAAGAGCTT 15401	
RESULT 3			
ABA01441			
ID	ABA01441	standard; DNA; 17468 BP.	
XX	ABA01441;		
XX	AC		
XX	21-FEB-2002	(first entry)	
XX	DE	Streptococcus thermophilus eps3 operon #2.	
XX	KW	Exopolysaccharide; lactic acid bacterium; eps; fermented food product;	
XX	OS	ds.	
XX	PN	Streptococcus thermophilus.	
XX	MO200179500-A2.		
XX	PD	25-OCT-2001.	

PF 18-APR-2001; 2001MO-FR001199.
PR 18-APR-2000; 2000FR-00004972.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA (DANO-) CIE DANONE SA GERAVALS.
PA (RHOD) RHODIA CHIM.
XX
PI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P,
DR WPI; 2002-017616/02.
XX
XX New nucleic acid fragments containing exopolysaccharide operon, useful
PT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
PI
XX
PS Claim 9; Page 75-80; 144pp; French.
XX
CC The present sequence is an eps operon from *Streptococcus thermophilus*.
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC synthesis. The operon is useful for producing chimeric eps operons, for
CC optimizing production of EPS in lactic acid bacteria. EPS impart texture,
CC mouth feel and rheological properties to fermented food products (e.g.
CC yoghurt). They function as thickeners, to provide free-flowing and creamy
CC texture, and may also have biological activities beneficial to health
XX
XX
SO Sequence 17468 BP; 5871 A; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;

Query Match 13.7%; Score 955.6; DB 6; Length 17468;
Matches 1484; Conservatve 65.8%; Pred. No. 2.1e-119;
Mismatches 729; Indels 41; Gaps 5

OY	1	ATGCGCAAAACGAAATTGGCATTTATGTATGATAGACAGTGGCAATTTCTGCAATCTTTAA	60
Db	6294	ACCGGCAATTTAGACTGGCGCTTTTAGAATGATTTGGGGGTGTGTGGCTGTCTCCATTA	6353
OY	61	CAAGTCATATACCAATGCTGATTTAAATCGTTCGAAATTTTATATCAATATGATGTTC	120
Db	6354	CAACCATTTGCCCACTCAACAAAGATGATGATGAGATGGACCGTATTGATTTATGGCCCTGC	6413
OY	121	ATTATTTTGGCATTTTATTTATCTGTATCCGATGGAATTTGATATAGAGTAACTGA	180
Db	6414	ATTTTGTGGCTTTTATTTCTCAGGATGTCATCGAATTTTGAGAGTGTAGAGGCTCACTTA	6473
OY	181	TAGAGTTGAAAAAACAATTAACATATGATATATATTTGCAATTTTCTTAGCGGAGAT	240
Db	6474	TAGAGTTGAAACGGGCTGCTAGATGATGTTTGTGATTTTGTGATTTACTTCTCATTTCAATTTT	6533
OY	241	CATTTTGTGGAGAAATATTTTCGCACTTTCAAGACGTGATGCCGTATTTTCAATTTAA	300
Db	6534	CTTTTATATGGAGATATGTTTATTCATTTACGCACGCGGTGTGGCATTTATTTACATTTG	6593
OY	301	TAAACTTCGTTTGGATATACCTATTTAAACGTAATTAATTAAGCAATTTAAGATAGCTTTC	360
Db	6594	TTACCTTATATGATGTTATATATATACATGTCGACTGTCGTAGATACAAATACCTCTTCT	6653
OY	361	TATTTTGCACATCTATCAAAAAAAGACGATTTCAATTTCAACGCGCTGAACGATGGGAAA	420
Db	6654	TTATGACTGCTGACCAACAGAGAAATCTTTAGTCATTTACCACTATGAGAGATTCGCAT	6713
OY	421	ATATGCAAGTTTATTTGGAATCACAATAAACAATTCAAAAAAATCTTTGSCATTTGATAG	480
Db	6714	TAAATGAGAGGGGCTCTTTGGAATGCGATCAACTCTTACCAAAATATTTGGCGGGATTTGTAG	6773
OY	481	TTTTTAGTACGAATAATGATTTAAATTTAATTTATCATTTACCGCTCTATTTATTTCTGTGAAG	540
Db	6774	TCATTTGG---AGACGGTATGTGGCTTTTCCAGAGGAGATTCACATATGTTCTTTTGATTT	6830
OY	541	AAGCTATAGATTTTCAACAAGGAGGTGTGCACACGTTTATATAATCTACCAAGTG	600
Db	6831	ATGCATATAGATTTTGCACCCATGATATGTGTGCACATGTGTATTATCAACTTACCGAGTG	6890
OY	601	AGTTTATAGCGTAAAGCAATTCGTTTCAGATTTTGAGTTGTATGATTTAGATGTAACG	660

Db	6891	AACATTAACGATCTCAAAACATCTGTTTCCGATTTTGAAGTCATGGGATTTGATGTCGAGTG	6950
OY	661	TTGATATTAATTCATTCGCTTTTACTGCGTTGAAAACAAAAAATCCAACTGCTAGTGTG	720
Db	6951	TAGATATTAACACATCTGATTTTAGGGCTTTAACAAATATAAAAAATCAAAACAGGTGGAG	7010
OY	721	ACCAATAGCATTTGTAACCTTTTCCCAAAATTTTATAGCCTAGTCATATCAATGAAAC	780
Db	7011	ACCAATAGCATCTGTAACCTTTTAATCCCAATTAACAAACATACGCAATCTTTCTAAAGC	7070
OY	781	GACTTTGGATATCTCGGAGCGGTAGTCGGGTAAATATTTGGTGATATGTTCTCAATTT	840
Db	7071	GCAATGGGATATCTTTGGGGCTTTGATTGGTCTTAAATTTGGTGTGCTGGGATTTG	7130
OY	841	TGTTAGTTCCAAATTAATTCGTAGAGATGGTGAGCCGGATATTTTGTCTCAGAAAAGTTTG	900
Db	7131	TCTTAGCCCCATCATTTGCGAANAATGAGAGACCGGCATTTTGTTCAGAAAAGTAGAG	7190
OY	901	GACAGATGAGCGCATTTATTAATTATCAAGTTTCATGATGATGATATGTTGATGCTGAGG	960
Db	7191	GGAATAACGCGACGTATCTTTAAATTTTAAATTCGTTCTATCTACACTGATGCGGAGG	7250
OY	961	AGCCCAAAAAAGACTTGCTCAGCCAAAAACGATGCAAGGCTGGGTATGTTTAAAAATGG	1020
Db	7251	A-----AAGAAAGAATTGAAGGACCAAAATCAGATCA-----CGTGGAAATGTTCAAGAT	7301
OY	1021	GAATAACGATCTTGAATTAATCTCCAAATTTGACATTTTCATACGCAAAAAACAAGTTTACGC	1080
Db	7302	GAATACGATCCACGATACACTCCGATCGGTGTTTTATTTG-AAAAACGATCTGGAGC	7260
OY	1081	AGTTACCAAGTTTATATATGTTTAAATGGCGATATGATGCTATGTTGGTATCACGTCAC	1140
Db	7361	AACTCCCTCAGTTTTCACAATATGTTGGCGCGATATGATGATGCTTTGTTGGACATCGCCAC	7420
OY	1141	CTACAGTTGATGAATTTGAAAAATTAATCTCGGTCAAAAAGACGATGAGTTTAAAC	1200
Db	7421	CAACTGTGATGAATTTAAAAAATATACCAAGTCAAAAACGTCGCTCAGCTTCAAGC	7480
OY	1201	CAGGATTTACAGGCTCTCGGACGAGTTAGTGTGCTAGTAAATATACAGACTTCGACGACG	1260
Db	7481	CAGGAATTAACAGTCTTATGCGACGAGCAATGTGCTGATGAATTAATTAACCACTTTGATGAGG	7540
OY	1261	TAGTTCCGTTGACTTAACATTAATTAATTTGATGCTGCTCAGATTTTAAATTT	1320
Db	7541	TGCTAAAGACTCGATGTGATATATCATTAATCTGATCATTTGGTCGATATCAAGATAT	7600
OY	1321	TATTTAAACAGTGAAGTGTATGTTGTAAGAGGGGAAGTAAGT-----	1365
Db	7601	TGTTGAAGCATATGTGTGCTTTTAAAAAAGGGAAGTATGATGAGCCCTCTTTCA	7660
OY	1366	-----AAAGATATGAAAGTTGTGTCGCTTTCAGGGGACATTTG	1412
Db	7661	TTCTTGATTTGGAGAAAAAATGAAAAATTTGTTAGTAGGTCATCTGAGGGCATTTA	7720
OY	1413	ACTCACTGTATTTGTTAAAAACGTTTGAAGAAAGAAACGTTTTGGGTAACTTT	1472
Db	7721	GCACATTTAAATATGTTAAAAACATTCGGAAGATGAAGAACGTTTTGGGTAACTTT	7780
OY	1473	GATTAAGAGATGCAAGAGTCTTTGGAAGATGAAAAAATGATATCCATGTTAATTTTCCA	1532
Db	7781	GATTAAGAGATGCAAGAGATCTTTAAAAAGATGAAAAAATGATATCTTTGTCACCTTTCT	7840
OY	1533	ACAAATCGCAATCTCATTAATTTAGGAAAAATACCTTTCTTAGCTTTCAAAATTTTACGT	1592
Db	7841	ACAAATCGAAACTTTAAAAATTTAATTAATAATTAATCTTTCTTAGCTTTTAAAGTATTA	7900
OY	1593	GATGAGAAAACGATGTTATTAATTTCAATCTGTCGCGCGCTGTGCTCCCTCTTTTAC	1652
Db	7901	AAAGAAAACCTGATGTCATTAATTTCAATCTGAGAGTGGGTATGCGCTTCATCTTTTAC	7960
OY	1653	ATCGGAAAACATTTGGAGCAAGACGATTTATTTGAAGTATTTGATCGAGTTAATAA	1712
Db	7961	ATTGGTAACTATTTGGAGCTAAGACGTTTATATTTGAAGTCTTTGATGAGATTAATAA	8020

QY	1713	TCACATCTAACTGGAAAACTAGTTTATCCGGTAAACAGATATTTTATTTGTTCACTGGGAA	1772
Db	8021	CCAACAGTTACAGAAATAATAGTGTACCTTGTCAGGATTAAGTTCATTGTTCATATGGGAA	8080
QY	1773	GAATGGAAGAGGTATATCTTAATCTATTATCACTTGGGGAGTATTTTATATGATTTTGG	1832
Db	8081	GAGATGTAAAAAGTTTATCCAAAAGCTATTAATTTTAAAGGAGTATCTTTAATGATTTTGG	8140
QY	1833	TAAAGTAGGAACTCATGAAACAACAGTTTATCGATTGATTAATAAAGATTTGATTTATGGA	1892
Db	8141	TACAGTAGGAAACACAGAACACCTTTTAATGCTTATATTCAGGAAGTGGATCATTTAG	8200
QY	1893	AAAAAATGGAAGTATACCGACGAAATTTATTTCAACAGATATTTCTGACTATATTC	1952
Db	8201	TTGAAACAGAGGTATTAAGGAAGAGTCTTATATTCAAACGATATTTCTATACGAAC	8260
QY	1953	CAGATATTTGCAGATATTAATAAATTTCTCAGTTACAAAGAAATGGAACAATATATTACA	2012
Db	8261	CTAAATTTTGCAGGTGCTCTGTTTGAATGTTTGAATCAAAATAGTAAATTTATGCAA	8320
QY	2013	AATCAGAAAGTATTTTGGACGAGAGCCCGCTACTTTATGATTAATTCATTAATCCAAAG	2072
Db	8321	AAAGTGATATTTATATCACTACGAGAGTCCAGGACTTTTATAGTGCCATTACAAATG	8380
QY	2073	GAATAAATCAATTTATTTGTTTCTTACAGAAAAAAGTATGTGAACATGTAAATGATCATC	2132
Db	8381	GCAAAAACCAATTTGTTGTGCTCTAGACAGAAAAAGTTTGGTGAACATGTCAATGACATC	8440
QY	2133	AAGTAGAGTTTGTAGAAGAAATTTTACAGATTAATATTTTATTTATAGAAAAATATAG	2192
Db	8441	AAGTAGATTTTGTCAAGAAATGTACAAAAAAGAAATGGAACTATCATCTGTGTTGAAGATG	8500
QY	2193	ATGATTTGTTGAAAAAATTTATGAAGTTCTAA	2226
Db	8501	TAGATTAATTTGGGGAGACTATGAAAAATTACAA	8534
RESULT 4			
ADSI3200	ID	ADSI3200 standard; DNA; 16032 BP.	
XX	AC	ADSI3200;	
XX	DT	16-DEC-2004 (first entry)	
XX	DE	S. thermophilus CNM I-2980 polysaccharide synthesis-related operon DNA.	
XX	XX	lactic acid bacterium; food; meat; cereal; dairy;	
KM	polysaccharide synthesis; operon; ds.		
OS	Streptococcus thermophilus.		
Key	Location/Qualifiers		
FT	CDS	342..1802	
FT		/*tag= a	
FT		/product= "Transcriptional regulator eps13A protein"	
FT		/note= "The corresponding protein sequence is not shown	
FT		shown within the specification"	
FT	CDS	1803..2534	
FT		/*tag= b	
FT		/product= "Polysaccharide polymerisation/export-related	
FT		eps13B protein"	
FT		/note= "The corresponding protein sequence is not shown	
FT		shown within the specification. The CDS has a prokaryotic	
FT		val start."	
FT	CDS	2543..3235	
FT		/*tag= c	
FT		/product= "Polysaccharide polymerisation/export-related	
FT		eps13C protein"	
FT		/note= "The corresponding protein sequence is not shown	
FT		shown within the specification"	
FT	CDS	3245..3985	


```
FT      /*tag= d
FT      /product= "Polysaccharide polymerisation/export-related
FT      eps13d protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      4042..5409
FT      CDS
FT      /*tag= e
FT      /product= "Undecaprenyl-phosphate glycosyltransferase
FT      eps13e protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification. The CDS has a prokaryotic
FT      Val start."
FT      5611..6195
FT      CDS
FT      /*tag= f
FT      /product= "Undecaprenyl-phosphate glycosyltransferase
FT      eps13f protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification. The CDS has a prokaryotic
FT      Val start."
FT      6251..6634
FT      CDS
FT      /*tag= g
FT      /product= "Undecaprenyl-phosphate glycosyltransferase
FT      eps13g protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      6643..7092
FT      CDS
FT      /*tag= h
FT      /product= "Beta-1,4-galactosyltransferase eps13h protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      7092..7607
FT      CDS
FT      /*tag= i
FT      /product= "Beta-1,4-galactosyltransferase eps13i protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      7597..8493
FT      CDS
FT      /*tag= j
FT      /product= "Rhamnosyltransferase eps13j protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification. The CDS has a prokaryotic
FT      Val start."
FT      8763..9797
FT      CDS
FT      /*tag= k
FT      /product= "Glycosyltransferase eps13k protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      9827..10969
FT      CDS
FT      /*tag= l
FT      /product= "Repetitive unit polymerase eps13l protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      10984..11793
FT      CDS
FT      /*tag= m
FT      /product= "Repetitive unit polymerase eps13m protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      11844..12578
FT      CDS
FT      /*tag= n
FT      /product= "Glycosyltransferase eps13n protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification. No start codon."
FT      12633..13016
FT      CDS
FT      /*tag= o
FT      /product= "Glycosyltransferase eps13o protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      13049..14482
FT      CDS
FT      /*tag= p
FT      /product= "Transmembrane transporter eps13p protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification. The CDS has a prokaryotic
FT      Val start."
```

```
FT      CDS
FT      complement(14614..15870)
FT      /*tag= g
FT      /product= "Transposase IS1193 protein"
FT      /note= "The corresponding protein sequence is not shown
FT      shown within the specification"
FT      FR2852604-A1.
FT      PD
FT      24-SEP-2004.
FT      XX
FT      17-MAR-2003; 2003FR-00003242.
FT      XX
FT      17-MAR-2003; 2003FR-00003242.
FT      XX
FT      (RHOD ) RHODIA CHIM SAS.
FT      PA
FT      Horvath P, Manoury E, Huppert S, Fremaux C;
FT      PI
FT      WPI; 2004-671092/66.
FT      DR
FT      New strain of lactic acid bacteria, useful for preparation of foods and
FT      PT pharmaceuticals, especially fermented dairy products, contains genes
FT      PT involved in polysaccharide biosynthesis.
FT      PS
FT      Claim 5; SEQ ID NO 1, 35pp; French.
FT      XX
FT      The invention relates to a novel strain of a lactic acid bacterium that
FT      CC contains at least one of 8 specific nucleic acid sequences all of which
FT      CC are defined in the specification. The bacterium of the invention may be
FT      CC useful during food preparation, particularly that of beverages and meat,
FT      CC cereal, or dairy products e.g. fermented milk, yoghurt, 'matured' cream,
FT      CC dessert cream, cheese, soft cheese, cheese spreads, cottage cheese, milk-
FT      CC based drinks, dairy product retentates and baby milk, where the products
FT      CC are derived from an animal and/or plant. The current sequence is that of
FT      CC the Streptococcus thermophilus CNCM I-2980 polysaccharide synthesis-
FT      CC related operon DNA of the invention.
FT      XX
FT      SQ Sequence 16032 BP; 5191 A; 2323 C; 3188 G; 5330 T; 0 U; 0 Other;
FT      Query Match 12.7%; Score 889; DB 13; Length 16032;
FT      Best Local Similarity 66.8%; Pred. No. 1.0e-110;
FT      Matches 1344; Conservative 0; Mismatches 655; Indels 12; Gaps 5;
FT      QY 317 ATACCTTTAAGGTAATTTAAGCAGTTAAGAGTACGCTTCTATTTGCAATCTA 376
FT      DB 5592 ATACTTATCACTGCACGTGCGTAGATACAAATATCTATCTTATGACTGTGAACA 5651
FT      QY 377 TCAAAAAGACGATTTCTAATTAACAACGCGTGAACGAGGAAATATGCAAGTTTATT 436
FT      DB 5652 ACAGAGAAATACATTAATGATCATTACACACACTGAGAGATTGCGATTATGAGGGGCTCTT 5711
FT      QY 437 TGAATCACAATMAACAAATTCAAAAAATCTTGTCATGTAAGTTTAAAGTACAGAAAT 496
FT      DB 5712 TGAGTCAGATCACTTCTCTCAAAATTTTGGCAGGATATGATGATCATCGAAGTGTGA 5771
FT      QY 497 AGATAAATTAATTTATCATTTACCGGCTATTTATTCGTGGAAGAACTTATAGTTTTC 556
FT      DB 5772 AGTGGCGTTTCCAGAGGAGTCCCAATATTCCTTTT--GATGATGCGATGACTTTC 5828
FT      QY 557 AACAAAGGAGTGTGACACGCTTTTAAATCTCAACAGTGAGTTTAAAGCTAA 616
FT      DB 5829 GACTCATGAAGTTGTGACCATGTGTTTATCAACTTACGAGTGAACATTTACATCTCA 5888
FT      QY 617 GCAATTCGTTTCAGATTTTGAAGTTTGAAGTATGATGATGAAGCGTTGATTTAATTCATT 676
FT      DB 5889 ACATCTTGTTCCGATTTTGAATCAATGCGTATGATGATGATGATTTAACTTATT 5948
FT      QY 677 CGGTTTACTGCGTTGAAAAAACAACAAATCTCAACGCTGATGAGCATGCAATTGTAAAC 736
FT      DB 5949 TGAATTCAGGCGCTTTAAAAAATTAACAAATCAACAGTTGGAGACCATATGATCGTAC 6008
FT      QY 737 TTTTTCACAAATTTTATTAAGCTATGATCATATCATGATGAACGACTTTGATATACT 796
```

Db 6009 TTTTAATCCAAATTACTACAAACATAGCCAAATCTTTCTAAAGCCAGTGGATATCTT 6068
 QY 797 CGGAGCGGAGTCCGGGTAAATTAATTTGGTAAAGTTTCTTAATTTTGTAGTCCAAATAT 856
 Db 6069 TGGGGGTGGATGGCTCTTAATTTGTGGTCTGGTCGGGATTTGTCTTAGCTCTTAAT 6128
 QY 857 TCGTAGAGATGGTGG-ACCGGCTAATTTTGTCTGAGAAAGAGTTGACAGATGAGCGCA 915
 Db 6129 TCGGAAGAGATGAGGAACCTGCTAATTTTGTTCAGAAACGAGTAGGGAAGAAACGGTCTGA 6188
 QY 916 TAATTAATTTCAATCAAGTTTCATGATGATGTGTGATGTGCGAGGAGCCGCAAAAAGACT 975
 Db 6189 TCTTTAATTTTAAATATCCGTTCTATGATATGATGAGCAAAAAACGAAAAAAGAAC 6248
 QY 976 TGGTCAGCCAAAACCAATGACAGGGGTGGTATGTTTAAATGAGAAAAACGATCTAG 1035
 Db 6249 TTAGGCTCAAAAACCAAGATGCA-----GGGTGATATGTTTAAATGATATGACCAG 6303
 QY 1036 AATTACTCCAAATTTGACATTTCAACGCAAAAAACAGTTTGAAGAGTTACACAGATTTT 1095
 Db 6304 GATTAACCAATTTGGTCAAGTTTATCCG-AAAAACAGTCTGATGATTTGCCACAGATTTT 6362
 QY 1096 ATATGTTTAAATTTGGCGATATGAGTCTAGTGTGATGACGTCCACTACAGTTGATGAAT 1155
 Db 6363 ATATGTTACTAGTTGGTATGATGAGCTTGGTAGTACTGCTCCACGACGAGTGAAT 6422
 QY 1156 TTGAAAAATATATCTCCGTGCAAAAAGAGACGATTTGATTTAAACGAGGATTAACAGTCT 1215
 Db 6423 TTGAAAAATATATCAAGTCAAGTCAAAAACGACGACTTAATTTTAAACGGGGATTTACGGCT 6482
 QY 1216 TCTGCAAGTTAGTGTGCTAGTATATATCAAGACTTTCAGCAGAGTTCCGTTGGACT 1275
 Db 6483 TATGCAAGTTAGTGGGCGTAGCAATATACGACCTTGAAGAGTTAACTCGATG 6542
 QY 1276 TAGCATATCATTTGATTTGATCTATCTGTGCAATATTAATTTTATTAAGAAGTCA 1335
 Db 6543 TTGAATATCATATGATTAATCTGTCAATTTGTGCGATATCAAGATATCTTGAAGCTATCT 6602
 QY 1336 AAGTTGATTTGTGAGAGAGGAAGTAAAGTAA--ACTATATGAAAGTTTGTGGTGG 1393
 Db 6603 TCGTAGATTTTAAAGAGAGAGAGATGATATATCAATGAAAGTTTGTGGTAGG 6662
 QY 1394 TTCTTCAAGGGGACATTTGACTCATTTGATTTTGTAAAAACGTTTGAAGAGAAAGA 1453
 Db 6663 TTCTTCTGTGGGCAATTTGGCACTTGATATGCTTAACCTTTTGGATGAGACAGA 6722
 QY 1454 ACGTTTGTGGGTAACATTTGATTAAGAGATGCAAGAGTCTTTTGAAGATGAAAAAT 1513
 Db 6723 CCGCTTTTGGGTTACCTTTGATTAAGAGATGCAAGAGATTTTAAAGATGACAGTT 6782
 QY 1514 GTATCCATGTTACTTTCCAAACAAATGCAATCTGATTAATTTAGTGAATAATCTTCTT 1573
 Db 6783 TTATCCGTTGCTATTTCTCTAAACAGAAATTTTAAATTTATGTAAGAAATCTTTT 6842
 QY 1574 AGCTTTCAAAATTTTACGTATGAGAAACAGATGTTATTTTCACTGGTGGCGCGT 1633
 Db 6843 AGCACTTAATTTTAAAGAAAGAAAGACCTGAGTTATTTTCAACAGAGAGACGCGT 6902
 QY 1634 TGCCTGCCCTCTTTTACATGGAAGAACTATTTGAGCAAAAGCATTTATTTAGT 1693
 Db 6903 AGCAGTTCGGTTTATCTTGTGTAACCTGTTGAGCGAAACCGTTTATATAGAGT 6962
 QY 1694 ATTGATCGAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1753
 Db 6963 ATTGATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7022
 QY 1754 TTTTATGTTGAGTGGGAAAGAAATGAGAGATATATCTTAATCTAATTAATTAATTAATTAAT 1813
 Db 7023 ATTTATGTTGAGTGGGAGAGAGATTAATCTGTCTATCTCAAGATTAATTAATTAATTAATTAAT 7082
 QY 1814 TATTTTAAATGATTTTGTGAACAGTAGAACTCATGAACAAGTTTAAATGATGATTA 1873
 Db 7083 TATTTTAAATGATTTTGTGAACAGTAGAACTCATGAACAAGTTTAAATGATGATTAATTAATTAAT 7142

QY 1874 AAAGATTTGATTTATTTGAAAAAAATGGAATATTAACCGAGCAAAATATTTATGAAACA 1933
 Db 7143 AAGAGATTTGATCGTTTAAAAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7202
 QY 1934 GGAATTTTCGACTATATTTCCAGATATTTGCAAGTATTAATAAATTTCTCAGTTACAAAGAA 1993
 Db 7203 GGTATTTTCAACTTAAGAACTTCATACCTGTGACTGGAAGAAACCTTATCTCTTATTTCTGAG 7262
 QY 1994 ATGAAACATATATTTTACCAATCAAGATGATTTTGGCAGGAGGCCCGCTACTTTT 2053
 Db 7263 ATGCAACTATACATGACTCAGTATCATTAATTAACCATGATGATGATGATGATGATGATGATGAT 7322
 QY 2054 ATGAATTCATTAATCCAAAGAAAAAACAATTTATTTTCTTAAGCAAAAGATATGAT 2113
 Db 7323 ATGGAGCAATTTGCCAAAGCAAAACCAATTTGTTTCCAAAGCAAAAGATATGATGAT 7382
 QY 2114 GAACATGTAATGATCATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2173
 Db 7383 GAACATGTAATGATCATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 7442
 QY 2174 TTTATTTATGAATAATATAGATATTTGTTGAAAAAATTTTGAAGTTTCTAAGCAACT 2233
 Db 7443 ATTTATTTGTTGAGATATATTAATTTCTTCACTGAGAAATTTTATTAAGAAATTAATGAT 7502
 QY 2234 AACTTTACATCAATATATATTTTGTGAAAGTTTAAACAAATTAATTTGATTAATTT 2293
 Db 7503 ATTTGCCAAAGCAGAGTCAATTTTACCTTAAGATGATCAAGTTTATTTTAAATTTCAAA 7562
 QY 2294 AATGAGATCAAGAAATGATTAATTAATAA 2324
 Db 7563 GAGATATGATGATCACTTTGAGAGAGAGA 7593

RESULT 5
 ABLN71527_11/c
 Continuation (12 of 22) of ABLN71527 from base 1100001 (Streptococcus polynucleotide SEQ
 Contination split into 22 fragments Locus ABLN71527 Accession ABLN71527

WP	Fragment Name	Begin	End
WP	ABL71527_00	1	110000
WP	ABL71527_01	100001	210000
WP	ABL71527_02	200001	310000
WP	ABL71527_03	300001	410000
WP	ABL71527_04	400001	510000
WP	ABL71527_05	500001	610000
WP	ABL71527_06	600001	710000
WP	ABL71527_07	700001	810000
WP	ABL71527_08	800001	910000
WP	ABL71527_09	900001	1010000
WP	ABL71527_10	1000001	1110000
WP	ABL71527_11	1100001	1210000
WP	ABL71527_12	1200001	1310000
WP	ABL71527_13	1300001	1410000
WP	ABL71527_14	1400001	1510000
WP	ABL71527_15	1500001	1610000
WP	ABL71527_16	1600001	1710000
WP	ABL71527_17	1700001	1810000
WP	ABL71527_18	1800001	1910000
WP	ABL71527_19	1900001	2010000
WP	ABL71527_20	2000001	2110000
WP	ABL71527_21	2100001	215561

Query Match 11.4%; Score 798.8; DB 6; Length 110000;
 Best Local Similarity 48.2%; Pred. No. 2.1e-98;
 Matches 3323; Conservative 0; Mismatches 3382; Indels 185; Gaps 31;

QY 16 TGGCATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75
 Db 46562 TAGGATGATTTCAAAACAGTTGTGGTTATTTTCTGCAAGTTTGAACATTAATTAATTA 46503
 QY 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCAATTAAGATGATGATGATGATGATGATGAT 132
 Db 46502 CTCCCAATTTTAAAGCAATTAAGATTAATTAATTTATTTGTTGTTATGATCAATTAATTTGCT 46443

OY	133	TTTTTAAATCTCGATGCGACGTGAATTTTGAGTAAAGGAATTCGTATGACCTTTGAAA	192
Db	46442	TTTATCTTTCGATTTTTTAAACAGACCTTTTGAGTCGTGCTATCTTGAAAGCTTTAAAA	46383
OY	193	AAACATTTAACTATAGTATTAATATTTTGCAAATTTTCTTAACGCGAGTATCAATTTTGCTGG	252
Db	46382	TGGRATTCGAATACAGCTTTTACTATATATTTTCATATCAAGTCAATATTTTATATTTTAA	46323
OY	253	AGAAATATTTTGCACCTTTCAGACGCGTGCCTGTAATTTCAATTAATATACTTCGTTT	312
Db	46322	AAAACCTCTTTTACACGACGACGACTTTCCTTTTTTACTTTTATATGCTATGAATTCGATTT	46263
OY	313	TGGRATTCCTATTTTAAAGTAATTTTAAAGCAGTTTAAAGATAGCTTTCTATTTTGCACAA	372
Db	46262	TATATATATCTATTAATTCATTTTTTAAAAATATATTCGAAATATTTCTTAACGTAACTTTT	46203
OY	373	TCTATCAAAAAAAGCGATTCCTATTTTCAACCGCTGAACGATAGGAAAAATATGCAAGTTT	432
Db	46202	CACGAGATACCAAGATTGTTTGATTAACGAATAAAGATCTTTATACAAAATACCTTTA	46143
OY	433	TATTTGATATCAATTAACAAATTTCAAAAAAAATCTTGTCAGTTGTAGTTTATAGTACAG	492
Db	46142	GGAATTAAT---ACGACCAATATATATATTCGTGTCGTATCTTGGAATTCCTCGAAAAAG	46082
OY	493	AAATAGATAAATTAATTTATTTATTCATTCACCGCTCTATTTATCTGTGGAAGACGTATAGAT	552
Db	46085	ATTGTTATGATTTGAACATTAATCTGTTTAAAGATAATTAACAAAGATGCTCTTACTTCAG	46028
OY	553	TTTCAACAGGGAAGTGTGTGACACGCTCTTATTAATTCACAGATGAGTTTATGAGG	612
Db	46025	AGTTTAACCTGTTTAACCTGTGATTAACCTTTTATTAACAATCCATTAAGTAATTTTGGTA	45966
OY	613	TAAAGCAATTC-----GTTTCAGATTTTGAGTTGTAGTAATTAAGTAAAGCTTG	663
Db	45965	AATACCAATACAAAGATATTTATTAATGACATGGAAGCAATGGAGATGTTGCAATGTTA	45906
OY	664	ATATTAATTCATTCGCTGTTTAACTGCGTTGAAAACAAAAAATCCAACTGCTAGTGACC	723
Db	45905	ATGTAGAGGCACTTAGCTTGAATTAATTAAGAGAAAAAGCAATCCAACTTTTGAAGAT	45846
OY	724	ATAGCATTTGTACTTTTCCCAAAATTTTATTAAGCTTACTATATCATGATGAACGAC	783
Db	45845	ATAGTGTTATTAACATATTTCTATGAAATTCATTAATTAATGTCACCTTAATAGCAAAACGAT	45786
OY	784	TTTTGGATATATCTGGAAGCGGTACTCGGGTTAATTAATTTGGTATAGTTTCTAATTTGT	843
Db	45785	TTTTGGATATACGGGGGTCTATATATAGTTTGCTCATATATGGCAATGTGGCAATTTTTC	45726
OY	844	TAGTTCCAAATATTCGTAGAGATGAGTGAACCGGCTATTTTGTCTCAAAACGATGTGAC	903
Db	45725	TAGTTTCCCAAAATCAGAAAAAGTGTGGAACGGCTATCTTTTCTCAAAATAGATAGTTC	45666
OY	904	AGAAATGACGCAATTTTACATTTCTACAAAGTTTCAGTGAATGATGATGTTGATCTGAGAC	963
Db	45665	GTAATGTAGATTTTATAGATTCATTAATTTACATCAATGCAAGTATGAGTACGAACAA	45606
OY	964	GCAAAAAAAGATTCGCTCAGCCAAAACCAAGATGCAAGGTGGGTAATGTTTAAAAATGGGA	1023
Db	45605	TTTAAGAAAGATTTATTTAGTTTCACATTAATATGACAGGG-----CTAATGTTTAAAGTTAGA	45551
OY	1024	AAACGATCTCTGAATTAATCTCAATTTGACATTTCTACGCAAAAAACAAGTTTGAAGAT	1083
Db	45550	CGATGATCTCTGAATTAATCTAATAATGGAATAATTTATTCG-AAAAACAAGATATGATGAT	45492
OY	1084	TACCAAGATTTTATATGTTTTTAAATTTGCGATATAGCTTAGTTGGTACAGTCCACCTTA	1143
Db	45491	TGCTCAATTCCTAATAATGTTTTTAAAGGTATATGATGATTAAGTAAACACGCCCTCCCA	45432
OY	1144	CAGTTGATGATTTGAAAAAATATCTCTGTGTCAAAAGACGATGACGTTTAAACG	1203
Db	45431	CAGTTGATGATTAAGAAAAAGTATATTTCAACGCAAGGACGACGCTTAAGTTTAAACCG	45372

QY	1204	GGATTACAGGCTCTGGCAGGTAAGTGGTGGTGAATATCAAGACTTGCAGACGTAG	1263
Db	4531	GAATCACTGGTTTGCCAAATATCTGTAAGAAATATATTAATCTGATTTGATGAAATCG	4531
QY	1264	TTCCGTTGACATTAGCATACATTGATTAATTGACATATCTGGTCAGATATTTAAATTTAT	1323
Db	4531	TAAAGTTAAGATTGTCATATATATCAATGATGATGATTTGGTCAAGATATTAAGATTATTC	4525
QY	1324	TAAAGACAGTGAAGATTGTATTTGTTGAGAGAGGAGTAAGTAAAGTAT-----	1373
Db	4525	TCTTAACATTAAGGTATGTTTTACTCGGGAACAGAGCTAATGTTAAAGTTAAGTTTGAAAG	4519
QY	1374	-----ATGAAATTTGTTTGGTCGGTCTTCAAGGGGACATTTGACATCTGATATT	1426
Db	4519	GAATATATATGAAATTTGTCGTGGTTGAGTGAAGTGGTGCATCTGACACACTTGAACT	4513
QY	1427	GTTAAACCCGTTTGAAGAGAAAGAAACCTTTTGGGTAACTTGAATTAAGAGATGC	1486
Db	4513	TTTGAACCCATTTGGGAAAAAGAAAGATGGTTTGGTAACTTTGATTAAGAAATGC	4507
QY	1487	AAGAAGCTTTTGAAGATGAAAAAATGATTCATGTTACTTTCGAACAAATGCAATCT	1546
Db	4507	TAGAGATATCTAAGAGAGAGATGATATATCATGTGCTTTCGAACAAACCGTATGT	4501
QY	1547	CATTATTTAGTGAATAATCTTTCTTAGCTTCCAAATTTTACGTGATGAGAAACCA	1606
Db	4501	CAAAACCTGGTAAATAATCTATTTCTAGCTTTTAAAGTCTTGAAGAAACCAACGA	4495
QY	1607	TGTTATTTATTTCAATCTGATCGGCGCGTGGTCCCTCTTTTACATCGGAAACATTT	1666
Db	4495	TGTTATCATATCATCTGGTCCGCTGTAGCAATCACTTTTATATTTGGTAAATTTAT	4489
QY	1667	TGAGCAAAAGCGATTTATATTTGAAGTATTTGATCGAGTTAATTAATCTACATTAATCG	1726
Db	4489	TGCTTTAAGACCGTTTACATAGAGGTTTTCGAACAGATGAATTAACAACTTTGACAG	4483
QY	1727	AAAACATGTTATTCCTGTAACAGATATTTTATTTGTTCACTGGGAAAGAAATGAAGAT	1786
Db	4483	AAAATGTGATCTCTGTAAACAGATTAATTTATTTGTTCACTGGGAAAGAAATGAAGAT	4477
QY	1787	ATATTCCTAATCTATTAACCTTGGGGAGATTTTATATGATTTTGTAAACATAGAGAACT	1846
Db	4477	TTATCTTAAGGCATTTATTTAGAGAGATTTTATATGATTTTGTCAACATGGGAC	4471
QY	1847	CATGAACACAGTTTATTCATTCATTAAGAGATTTTATTTGTAAGAAAAATGGAAGT	1906
Db	4471	CATGAACACAGTTTCAACCGCTCTTATTAAGAGATTTATTAAGAGAGAGTGC	4465
QY	1907	ATTAACGACGAATATTTATTTCAACAGATATTTGACTATATTTCAAGATATTTGCAAG	1966
Db	4465	ATTGATCAAGAGGTGTCATTTCAACCGGTTTACTAGACTTTGAACCTCAGATTTGTCAG	4459
QY	1967	TATTAATAATTTCTCAGTTACAAAGAAATGGAACATATATTAACAAATCAGAAAGTAT	2026
Db	4459	TGCTCAAAATTTCTCATATGATGATATGAACCTTTACATGAAGAAAGCTGAGATTTGT	4453
QY	2027	ATTTCACAGGAGGCCCGCTACCTTTTATGATATTCATATTCAAAGAAAAACAATTA	2086
Db	4453	ATCACACATGGCCGCCACGACGTTTATGATATGCACTTTCTTAAGAAAAAAAACATTT	4447
QY	2087	TTGTTTCTTGAACAAAAAGTATGTTGAACATGATTAATGATCATCAAGTAGTTGTGA	2146
Db	4447	GTGGTTCTTGAACAAAGACGTTTGGAGACATGTAATTAATCATCAGGTGACTTTGTT	4441
QY	2147	AGAGAAATTTTACAAGATATATATTTTATTTATAGAAAAATATAGATGTTGTTGAA	2206
Db	4441	AATTAAGTAAACCAATGATATATTTGATATGTTGTGTGATATTAAGGTTACAAAT	4435
QY	2207	AAATTTATTAAGTTCTAAGCAAACTAATCTTACATCAAAATATATATTT-----	2263
Db	4435	GTAGTCTATAGAGGAATGATGAATCGTCCGTTTATTAAGAACTATATGTAGTAAATTTTAT	4429
QY	2264	GAAGAATTAACAAATATGTTGAAAAATTTATGAGATTCAGAAAAATGATATATAAAA	2323

Db 44291 GAAAGATTAAAGTATATTAAGAGGTGTGCGATGAAATCAATAAATCTTTAT 44232
Qy 2324 AGATGATTTTGAATATGAGCTTATCATATTTTCTCGATTTTACGAGAGGGATAC 2383
Db 44231 TTATTTGCAATATTTTATGTAATTTTAAATACGCGGTTTACGAGAGAAATCTC 44172
Qy 2384 AGATTTATCATCTTCTCTCAGGAGAAATGCACACATTTAGTTCCTTCAAGATACCTGTAT 2443
Db 44171 AGCTTACAAATATGTC-----ATGTTAGTTGCA 44143
Qy 2444 AATTTATTTAAATTTCTCAGATTTATATGTTGAATTTACAAAGATGACAAATAT 2503
Db 44142 TTTTACGTGTGGAATTAATTTTATTTATGATAGCTTTATTTTGAAGAAAGAAATCTCG 44083
Qy 2504 AAAAGAAATAGATATATGACGAGTAAATGTTACAGATTTATTCCTAATATATCAGAA 2563
Db 44082 TGATCATCTTTTATTTATTTATTCGACACATTTTGAATTTATCT----- 44038
Qy 2564 AAAACATTTGATATGATGATCGTTTGAATTTTATGAAGATGATCGACTTTTGAATAC 2623
Db 44037 ----TTGTTCAATAGGTTACTTTATATTTAATTTTCTTACGATTTAAAGAC 43982
Qy 2624 TATTTACAAAGATGTTGTTTATATGATATGAAATTAACATGCTTAAAGATTAAGTTTG 2683
Db 43981 ATCTCTTAAATAAAGCTTCTCTATATATATGATTCGCGTATTTGGAGTTCTATTA 43922
Qy 2684 GTTCTAATTTGGGTTTGGCTTCCACATGATTTTGTGCAATCTTTATCAATGAAGAAAG 2743
Db 43921 AATCAAAATTTTGTGAATTTAGATTTAATGAAATTAAGATATGCAATTTTATAGGAT 43862
Qy 2744 AAACAGCTTATTTATTTAAGTATCTAAATGTCAGATGACATTTATTAACACACATTA 2803
Db 43861 GGCATATTTATTTCTGAGAAATGA-CTTAGTTTGTGTCATCTTATTTATCATATTT 43803
Qy 2804 TAGAAAAATATGAAATTTTCAATATGATTTCTAATATGAA--ATTATGATATATTA 2860
Db 43802 TTTTGTCTATCATATTTTCTGTATATGATCTCAATATTAACGACTAAAGCTGTGTG 43743
Qy 2861 AGTGAAGAAAAATCAACATCTTCTCCATGTCTTTTCAAGATGATCTATGATGAAATTCG 2920
Db 43742 GATGTTTTTATTTTAAATTAATTTATTTATGTAACCAATATACCTTTTCAAGACAGG 43683
Qy 2921 TAAATGCAAGAAATTTAGTTTTTTTATTTTCTAGAAAGTTAAATAATGAATATATCTA 2980
Db 43682 GTATATATGCTAATTTTATTTATGTACTATTTATGTGACAAAGAAATGCTTAATTA 43623
Qy 2981 AA----TTTAAAGAAATTTATTTACTTAAATAAATAATAGTTTGTGTGAGATATGA 3036
Db 43622 AAGAGATTTATGAAATTTAGCACCCATGACCAATTTTATTTATGATATTTACCTTTT 43563
Qy 3037 TGTTTAAATTTATTAATATGACCCGGAATTTTATTTTAAAGTACTTGTGTGATTA 3096
Db 43562 GAGTTCTACAAATTTTATTTCAATTTTGTTCAAATAATGAGATGTTCTTTTAAACAG 43503
Qy 3097 TTTTATTTCCAGCAAAAGATGATTTTATTTATTTTATGAATTTATTTATTTATTC 3156
Db 43502 TAGATTAACCTATGCTCATTTACACTGTGATGTTTAACTCTTTTGAAGATGTT 43443
Qy 3157 ATATTAATTTTGAATACTAGCTAATTTATTAATAAATAATGAATTTATTTTATTTAT 3216
Db 43442 TAAAGAAACAAGTGTCTATTTGATATATAGCTATGTAATGATATGATATGATGT 43383
Qy 3217 GGTCTATATTTATTTTGTGTTCACTAGTACACAGTATGTTTGTGAATAATTTTGAA 3276
Db 43382 AGTACTATACATGTTTGTATGATATATCTATATATCTATATGTAAGAA--TAA 43330
Qy 3277 GATTTATGAGATTTTACTGCTCCATTAATTTGATTTATGCAATATGATATTTATTT 3336
Db 43329 TCAATATGAACTCAACTACTCTATTTATATATGCTATATATTTTATCTGAAAGTT 43270
Qy 3337 TGTATTCATTTATTAATATTTGATTAATAAATAATTAATAATAGTATCTTTTATGTTTT 3396

Db 43269 TTTATCCAGTGTGTAATGAAATATTTAGTTGGCTAGTTTTTGTGTAATAATTTTGTGATG 43210
Qy 3397 TAGTTTTATTTAGATATATCTGATTTGTATTTATTTCAAAATGGAAGATTTGTATTTT 3456
Db 43209 GTATGAACCTATATAAAGAAATTTTACTATTTGTGAATAT--ATATGACATATTTGCTC 43152
Qy 3457 TAGACAGACACTTATAGACTATGCTATCTTTATTAACAGCGCTCAAAACAGTTGGTTG 3516
Db 43151 TGATATGCGAGAGGTAAGAAAGAAATGATATCTTAAGTTATATCATTTATTTGTTGTTT 43092
Qy 3517 GCTTATGAACTATGCTTCGTTAATATACATGATATATATGTTTCAATCCGTTATCT 3576
Db 43091 GGAGAAATTCCTTACCGATATTTTAAAGAAATATTAATAAATCTTGAGAGAACATATCT 43032
Qy 3577 TTGCACTATTAATAAATAATTAATGCAACAATTTTTTCTGTGCTTCTTTATATCCGA 3636
Db 43031 CCGGATATGAAATTTATGAAATGAAATGACATATTTATGATGT----- 42988
Qy 3637 TCTATTTAAGTGAATGCAATTTGGTATTTATTCCTTACCAATATTTATATATGCTTGT 3696
Db 42987 --TAGTAAATAATGTTTTATGAGAGACATATCTAAGAAATTTGCTTATGTTCT 42930
Qy 3697 TATGAGATATATAGG-TGGAATAATTTGCTGATTAATAAAGCTAATAGTAAATTTTGA 3755
Db 42929 GACTATGCAAGTTGATATTTATTTATCTTATGAGGGGTTCTATCTGATATCTGATGTG 42870
Qy 3756 ATACTACTATTTATTTTAAATCTGAATTTGCTTTTCAATGAATTTTGGCTGTTATAT 3815
Db 42869 GAGCTTTTAAAGTTTATGATCTTTTGAAGATTCATGAGTGTTCATGCAAGGAGATTT 42810
Qy 3816 TCTAGGAATCAAGTACAGACACTGATTTATTTATTTATCAAGGAATTTGATATAAGTA 3875
Db 42809 AGTTGATGATGAATAACAGATTTAATTTGCGCTGTTAAAGGACATCCTTTTAA 42750
Qy 3876 TTAGAAAAATATTTTATTTGATATGGAATATCCGAATATTTCAATTCGGGAATTTG 3935
Db 42749 TCAATATGCTATATATGACAAAGATATTTAATCTTATTAAGACATGTGATAGG 42690
Qy 3936 CTCGAAATCATTCAGGCTATATATCTTTTATTAATCAGAAATGTTGGTTGATTT 3995
Db 42689 GTTACACATATTTATTTGATTAACAGAGGCTTTAAGAT--AAGATATATTTCAAAAGA 42632
Qy 3996 TTAATGATGTTTTCTTTTATATGTTATTAATAAATAAAGTATAGACTTAATGGGAAACA 4055
Db 42631 TTGATGATTAACAATATATTCAGAAATATTTTAAATCAAGAAATTTATTAACAGGTA 42572
Qy 4056 GCACTATTTATTTTATGATCATTTAGCATTTTTCATATGATGAACAATATGATCCGAT 4115
Db 42571 AGGTGATTTGCTGACTAGTGTATCTATTTCTATATCATTTACGAGAAAGTTGAAAA 42512
Qy 4116 ATATATATATTTAGTACTATTTCTTTCTCAATAGTATTTGAAATATTAATTTTAA 4175
Db 42511 GTTCTTCATTTATTTCAATTTCTTAAAGATTTAGAGTAAAGGCTCAATATTTGATTTTAT 42452
Qy 4176 AAGATATGAGACAAAAATGAATGATTTAATTTCAATTTTGAATTTGAATTTATATGT 4235
Db 42451 TTGATATGTAATCTTATTAAGATCTCTAAGGTTTCTAAGTTTAAAGAAATATGTTATAT 42392
Qy 4236 CCAAGATATCTTGAATAATGATTAACAGTATTTAATTAACCAACATATACTAATTTAG 4295
Db 42391 TGTATATGAAGACAAAAATGTTATATTTTATTTTACCTTACATTTAAACAGAGGCTACGA 42332
Qy 4296 GGTATTTCTGTAATATGAGAAATGATGATATTTCTGAGAAATTTTGGCTTAAACTATAT 4355
Db 42331 TTATTCATATC-TGTATGATTTCTCTATGATATCGAATAATTAATAATTTTATATGTTGAT 42273
Qy 4356 GAAAGACATGAAAGATTAATATTTTACAAAGAAATTTAATGCGGCTTACAGAGATGCTG 4415
Db 42272 TGTGATGATGTTCTGAGAGACGATCTAAGAAATTTGTAAGTAATTAATAAAGAGAA 42213
Qy 4416 AATTTTGCACATGAAATGCAACAGTAAATATTTTGTCTTTGTGATTTCTGATGCTA 4475
Db 42212 TAAAGTTATGATTTGTATATCTGTA--TAAACGTAATGGCGTTAAGCATTTCAACCTTATA 42156

QY 4476 TATAGAGTTCAGATGTTCCAGAGATGATGATATATTAAGTATATGCGCAT 4535
| | | | |
Db 42155 TCTAGCAATGAGATATATGCAACCAAGTATATCATGTAATGCTAGATGATG 42096
| | | | |
QY 4536 AGCAGAGATGATTTTGTGTAGTACGAAAACGGGTATACAAAAGAAAAAGATG 4595
| | | | |
Db 42095 GTTATTTAGAGATGCGAGTTAAATTTTAAAGATTTAGAGCGCTAACGTTGCTTAA 42036
| | | | |
QY 4596 TAAATTTTCATGCTTACGAGAGAGAGAGAGCTGTAAAGAAATTTTTCAGAGATATAT 4655
| | | | |
Db 42035 TAGATATGTAAGCCT-----TGTTAATCCTCGATATTCATTAACCAAGGTAAATATG 41982
| | | | |
QY 4656 AGAAAATATGTTTGTGCAAGCTTTATTCACGAGATATTAATAAGATATAATTTCA 4715
| | | | |
Db 41981 GTTGAATCCCAAAATATTAAGATTAATTTCCGATTTAAATTAATTAATTCCTTAA 41922
| | | | |
QY 4716 AATTAATATAGAGATATGAGATGAGATTTGCTTTTAAATTTGAGAGCTTGAACAAATG 4775
| | | | |
Db 41921 AATTTGAAACTGTATGTTATTAATAA--TGCTTATTTAGATGATTTGCAATTTCTTGT 41864
| | | | |
QY 4776 AACACGTGATGATGATATGATGAGATATATTAATTAATGATGCTATGCAAGTTC 4835
| | | | |
Db 41863 TTGAAAGGGAGATTTCTATCAGAAATATATATATTAATTAATTAATTAATTAATTA 41804
| | | | |
QY 4836 GCTTATTAATCAGAAATTCCTATTAATTAATTAATTAATTAATTAATTAATTAATTA 4895
| | | | |
Db 41803 TACTTTTGTCCCAAAATAGAGAAATCTATTTG---TTTGACTACTTTAGAGATGAGT 41748
| | | | |
QY 4896 CCCCCCTTAAGTAAAAAGAGATTTAGTATTAATTAATTAATTAATTAATTAATTAATTA 4955
| | | | |
Db 41747 GACTTCATATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 41688
| | | | |
QY 4956 GGTAAATGTTTAAACAAAATGATTCACAGATTTGTTGATATGATGTTCTTGCAAT 5015
| | | | |
Db 41687 AAATTCATATATGATATGATGCTTTTCTTAATATATTTGATGCTGAGTGCTGAGC--T 41630
| | | | |
QY 5016 ATTAGAGCTTATTCGAAAAGAAATAGCTAGATATCCATTTATTAACGAAAAGATATTT 5075
| | | | |
Db 41629 ATTAAGATTAATAATTAAGAACATACAGCTTGAATAATGCTATTAAGGGGTAAATCTTACA 41570
| | | | |
QY 5076 ATCAAGAAACATTTAGTATGCTGTTATTTGATGCAAAATTTTCCGCTAACCTATATGAT 5135
| | | | |
Db 41569 CTAAAG-----AGTGAAGAGCGCTGGGTGGAATTTTGTGGGTTTATGCTATCTT 41519
| | | | |
QY 5136 GTTATATAGAAATTTCAAAAGCAGTACAGGTAAATTTGATTAATTAATTAATTAATTT 5195
| | | | |
Db 41518 TCGAAGGTACCAAGTTTAAAAAGAGCGAGTAAAGATGAATTAATTTCAATTAATAGTG 41459
| | | | |
QY 5196 CCGATTTATATAGTATTAATTTTAAAGTATGATTAAGATTAAGATTAATTAATTAATTA 5255
| | | | |
Db 41458 CCGTATATATATAGAGAAATATATTTAGGAGATGCTAGATAGTATTTCTTGAGCAAACT 41399
| | | | |
QY 5256 TATTAATAATTAAGAAATATATTTAGTATGATGCTGCTAGATGATTTCTGCTAAATA 5315
| | | | |
Db 41398 TATCAAAATTTAGAGATATCATATATGATGATGCTAGTATGCAACAGGGGATAT 41339
| | | | |
QY 5316 TCGAAGAAATATCAGAAAAAGATTAAGAGTAAAAATTTTTCACATAATCAATAGTGA 5375
| | | | |
Db 41338 TCGAAGAAATATTTTGTGAAGAAGAGGCAATTAATTTCTATCAAGAAAAAGAGGCG 41279
| | | | |
QY 5376 GTATCAAAATGCTAGAAATATCATGCAATTAACGAGATCACTGAATATTAATTTGTT 5435
| | | | |
Db 41278 CAATTCAGTTGACAGTAAACAGAGATCTTAAGGTATGACAGAGATGAGATTTGCTTTT 41219
| | | | |
QY 5436 GACTTCATGATGTTGTTGATAGATATGATGATTAATTAATTAATTAATTAATTAATTA 5495
| | | | |
Db 41218 GACAGATGATGATTTATCTTCTTATTCATATGAGGTATGATCAACATACAGAAAGCT 41159
| | | | |
QY 5496 AGTGAAGATGATTTATCTGTTGTTGATGCTATCTTTTCAAGAAATATTAATTAATTT 5555
| | | | |
Db 41158 ACCAATGCGCATATGTTTAACTAGATTTGTAATCTTCAACATATCTTCAATATAGT 41099
| | | | |

QY 5556 GAAGTGAATATCCAAATATGATTTTGAAGCAATTAATACCGTCAGAGATGAGAGAA 5615
| | | | |
Db 41098 ATTAATTCACAGATTTTAAAGAGATTAATCTTATATCTTTAGAGGTGCGCTGGAAGAA 41039
| | | | |
QY 5616 AAAATTTTATGATTTGATATTAATTAATTTTCTACTCCTGTTTGTAACTATAT 5675
| | | | |
Db 41038 ATGATATATGGAAGAACTTATGAGTTCGCCGTACCA-----AATTAATCCGAGAA 40984
| | | | |
QY 5676 AAGAAAGATATCATACAGATCTTTTCAAGAGATCAATGTTAGAGAAATTTACTT 5735
| | | | |
Db 40983 GTAAATTTATGATATCCATATCCAGAGAGAAATTTCAATGAATATGATACACTT 40924
| | | | |
QY 5736 TTTAATCTGATTAATTTAAAGAAATATGATAGATGATTAATTTGACTGAACATCTTAT 5795
| | | | |
Db 40923 TTAAGCTTATATCTTGCGCTTCAAAAATAGAGTTTGATGATTTGCTACGTCAGGTGCT 40864
| | | | |
QY 5796 TTTTATGAGAGATATCTTATGATCACTAAATCTTTTAAAGAGGTGTTTGTGCA 5855
| | | | |
Db 40863 ATTTTACGATTAATATGATCTATGAGCAAAATTCATGAGAGATGT--TGTATTTT 40805
| | | | |
QY 5856 TTGGAATTTTGCAGAAACAGATGATGTTGTTTAAAGCAATATATGAGAGATTTT 5915
| | | | |
Db 40804 GAGCAATTCAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40745
| | | | |
QY 5916 GAGTATCAATTTGTTAAGATTAATCTATGCTTGGCAATATTTTATTAATTAATTAATTA 5975
| | | | |
Db 40744 ATCTGCTGTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40685
| | | | |
QY 5976 TTTAATATCGAGAAACAGTATTTTGAACAATTTTAAATTTTAAATCTTTATTA 6035
| | | | |
Db 40684 GATTTTAAATTAATATGATACAGTTATTAATTAATTAATTAATTAATTAATTAATTA 40625
| | | | |
QY 6036 AAATATATTTTAACTTTTAAAGATCTTAACAAAATCTTTGCTTAAATTTTGT 6095
| | | | |
Db 40624 ACTATTTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40565
| | | | |
QY 6096 ATTAAGATTTTTCGAAACAAAGTTTAAATAATTAATTAATTAATTAATTAATTAATTA 6155
| | | | |
Db 40564 TCCCTATTAAGATTTTCACTATTTGTAAGAAAAATA-----TATATTTAGGATATC 40512
| | | | |
QY 6156 ATGATATCTATTAATTAATTTCTATTAATTTGATCTATATTAATTAATTAATTAATTA 6215
| | | | |
Db 40511 AAGGA-----ATGATTAAGTTAGATATATCCCAATTAATTAATTAATTAATTAATTA 40456
| | | | |
QY 6216 TCTAATATGATATGATGATTTGAATCAACACCTACAAACATATAGATTTCTGCTG 6275
| | | | |
Db 40455 AACGAGTATTAAGAAATGTTTATGCTCAAACTTATTAATTAATTAATTAATTAATTA 40396
| | | | |
QY 6276 AATGACGTATGATGATTAATTTGGAAGAAATTTGTTAGCATATGCGAAGAAAGATAT 6335
| | | | |
Db 40395 AACGATGCTCTACAGATTAATTTCCGGGATATCTGATATTAATTAATTAATTAATTA 40336
| | | | |
QY 6336 CCGATTCGTTATTTTAAAGAAAGAGAGCGGCGCTATCAGATGCCGCTAATTAATGAGATA 6395
| | | | |
Db 40335 CGTATATTTGTTTTCAGAAAATTAATGAGTTTATCAGATGCTGTAATTAATTAATTA 40276
| | | | |
QY 6396 AGTCGCGCCAGAGGTGATCTTAAGCTTTATTAATGATCAATGATTTTATTTATTTGAG 6455
| | | | |
Db 40275 AGTGAAGCTATCTGCGATTAATTTTATTAATTTGATTTGATGATGATTAATTAATTA 40216
| | | | |
QY 6456 TT---CATCCAGTTTACAGAGCAATGAGAGAGAGATGCCCTGTGCGAGTGT 6512
| | | | |
Db 40215 GATGCAATTCAGAGAAATGTTGAATTTTGGGAGAAATATTAATTAATTAATTAATTA 40156
| | | | |
QY 6513 GGTATG-----ATAGGTATATGCTTCGCGCATTTCTTAACAGAGAGCGCTTCT 6566
| | | | |
Db 40155 TGTATGTTGAAAAAGAGAAACAATATTAATTAATTTCTTTTAAAGATGAATGATA 40096
| | | | |
QY 6567 ACAATTCAGGCTTTCTGAGCGGAGAGATGTTTGTAAAGCTGCTAGAGGGGAGATGT 6626
| | | | |
Db 40095 GAAACTATTAATGTCACAGTACAGAGATTAACAAATTAATTAATTAATTAATTAATTA 40039
| | | | |
QY 6627 CATGCTTTGTGCTGCTGTATTAATCTATTAATTAATTAATTAATTAATTAATTAATTA 6686
| | | | |

```

Db      40038 GCTATTTTACTGTAGACATATAATATATATAAGAGAAATTTTCTACGCTGTC 39979
QY      6687 TTTGAAAAGGTTAAAGTTATGTAAGTGAATCTTACTATTCGCTGCTATAGTGA 6746
Db      39978 TATCCAGTAGGGAAGTTACACGAAGTGAATTTTAACTATTAATTAATTTATTTGAAAGCC 39919
QY      6747 GAAAAAGTTGCATAGTTAAGAGTGTCTGTACTATTATTTGTTGACCGAGAAATAGTATC 6806
Db      39918 AAAAATATATATTTCTTTTGAATATATATCTTATGCAATTCGATTTAGAGAAATAGCATA 39859
QY      6807 ACAACTTTCAGATGACGACCATCGCTTCCTCATTCGCTTACGTAATTTCA 6856
Db      39858 ATGACTGCTTCTCACAATATATTAAGGATACATGCTGTAGAAAGCACTTAA 39809

RESULT 6
ABK90550
ID      ABK90550 standard; DNA; 6865 BP.
XX
AC      ABK90550;
XX
DT      15-NOV-2002 (first entry)
XX
DE      Beta1,3-galactose transferase DNA #2.
XX
KM      Beta1,3-galactose transferase; gene; ds; galactose; N-acetylglucosamine;
KW      beta1,3-bond; galactose-containing glucoside; uridine-5'-diphosphate;
XX
OS      Streptococcus agalactiae.
XX
FH      Key
FT      CDS
          Location/Qualifiers
          ..1792
          /*tag= a
          /product= "Beta1,3-galactose transferase #2"
          1816..2262
          /*tag= b
          /product= "Beta1,3-galactose transferase #3"
          /partial
          /note= "No stop codon given"
          2265..2747
          /*tag= c
          /product= "Beta1,3-galactose transferase #4"
          2843..33979
          /*tag= d
          /product= "Beta1,3-galactose transferase #5"
          /partial
          /note= "No stop codon given"
          3982..4956
          /*tag= e
          /product= "Beta1,3-galactose transferase #6"
          5009..5950
          /*tag= f
          /product= "Beta1,3-galactose transferase #7"

CDS
FT      CDS
          JP2002199885-A.
XX
PD      16-JUL-2002.
XX
PF      05-JAN-2001; 2001JP-00000392.
XX
PR      05-JAN-2001; 2001JP-00000392.
XX
PA      (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX      WPI; 2002-612563/66.
DR      P-PSDB; ABG31703, ABG31704, ABG31705, ABG31706, ABG31707, ABG31708.
XX
XX      Beta1,3-galactose transferase and a DNA encoding the enzyme, a protein
PT      having beta1,3-galactose transferase activity, a transformant,
PT      preparation of a protein, and preparation of a galactose-containing
PT      glucoside.

```

```

XX      Example 1; Page 15-22; 24pp; Japanese.
XX
CC      The invention relates to a protein having beta1,3-galactose transferase
CC      activity, derived from a microbe having an activity of transferring
CC      galactose to N-acetylglucosamine by beta1,3-bond. The protein can be used
CC      in a method for the preparation of a protein having beta1,3-galactose
CC      transferase activity in which the above transformant is cultured in a
CC      medium to form and accumulate a protein having beta1,3-galactose
CC      transferase activity, and in a method for the preparation of a galactose-
CC      containing glucoside in which a culture liquid of the above transformant or
CC      a treated product of the culture liquid is used as an enzyme source, the
CC      enzyme source uridine-5'-diphosphate galactose and a receptor glucoside are
CC      made to be present in an aqueous medium, and a galactose-containing
CC      glucoside is formed and accumulated in the aqueous medium. The galactose-
CC      containing glucoside is used as a candidate for infection preventors. This
CC      sequence represents DNA encoding beta1,3-galactose transferase proteins
CC      of the invention
XX
SQ      Sequence 6865 BP; 2453 A; 847 C; 1185 G; 2380 T; 0 U; 0 Other;
XX
Query Match      10.4%; Score 727.8; DB 6; Length 6865;
Best Local Similarity 59.2%; Pred. No. 7,7e-89;
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;
QY      16 TGGCATTATTTGATATGATAGACGTTGCAATTTCTGCAATCTTAAACAATCATATACCA 75
Db      438 TAGCGATGATTCAAACCGTTGTTGTTATTTTCTGCAAGTTGACATTAACTTAATTA 497
QY      76 ATGCGTATTTAATGCG---TTCGGAATTTTATATCATATGATAGTTCATTAATTTTGCAT 132
Db      498 CTCACAACTTTAAMACCAATTAAGATTTATGTTTGTCTTATGATATCATTAATTTGTCT 557
QY      133 TTTTATATCTCGTATGCGCAGTTGATTTGAGTATAGATATCATATGATAGTTTGA 192
Db      558 TTTATCTTCTGATTTTACAGAGACTTTTGAGTGTGCGTATCTGGAAGTTTAAAA 617
QY      193 AAACATTTTACATATGATATATATTTGCAATTTTCTTACGCGAGATCATTTTGTGG 252
Db      618 TGGTATTTGAATACACCTTTACTATATTTTCAATCAAGTTCAATTAATTTTAAATTTCTA 677
QY      253 AGAATATTTTGGACATTTCAAGCGTGGCGCGGTATTTCAATTAATTAATCTGTT 312
Db      678 AAACCTCTTTTACACGACGCGCTTCTTTTATCTTTATGCTATGAAATTCGATTT 737
QY      313 TGGTATACCTATTTAAGTATATTTAAGACGTTTAAAGTACCTTTATTTTGCACAA 372
Db      738 TATTAATATCTATTTGAATTTCTTTTAAATATTTATGAAAATATTTCTTACGTAAGTTT 797
QY      373 TCTATCAAAAAAAGACATTTCTAATTACAACGCGTGAACGAGGAAATATGCAAGTTT 432
Db      798 CACGAGTACCAAAAGTTGTTTGAATACGAATTAAGATTTCTTATCAAAAATACCTTTA 857
QY      433 TATTTGATATCATTAACAATTCAAAAAAATCTGTGTGATGTGATTTTATGATGAT 492
Db      858 GGAATTAAT---ACGACCAATTAATTAATGCTGTCTGTATCTTGGAATTCCTCTGAAAAG 914
QY      493 AAATGATTAATTAATTAATTTATCATACCGCTATTAATTTCTGGAAGAAGCTATAGAT 552
Db      915 ATGTTATGATTTGAAACATTAATCTGTTAAGATATTAACAAGATGCTTAACTTACAG 974
QY      553 TTTCAACAAGGAAGTGTGACACGCTTTTATTAATCTTACCAAGTGAATTTTATGACG 612
Db      975 AGTTAACCGCTTAACGTTGATCAAGCTTTTATTAACATACCAATGAATTTATTTGGTA 1034
QY      613 TAAAGCAATTC-----GTTTCAATTTTGAAGTTTATGATATGATGAAGTGAAGT 663
Db      1035 AATACCAAAATTAAGATATTAATTAATGAATTAAGCAATTAAGGATGATTTGCAATGTA 1094
QY      664 ATATTAATGATTCGTTTACTGCGTTTGAAGAAACAAAAATTCACATCTGAGTATGAC 723
Db      1095 ATGTAGAGCACTTACGTTTGTATATATATAGAGAAAAGCAATTCACAACTTTGAAAGAT 1154

```

QY 724 ATAGCATGTAACTTTTCCACAAATTTTAAAGCTACTCATATCATGATGAACGAC 783
 DB 1155 ATAGGTATTTACATATTTCTATGAAATTTCTATTAATATAGTCACTTATGCAAAACGAT 1214
 QY 784 TTTTGGATTTACTCGGAGCGGTAGTCGGGTAAATATTTTGTGTATAGTTTCTATTTTGT 843
 DB 1215 TTTTGGATTTCAAGGGTGTCTATTAATAGTTTGTCTATGTGCAATGTGGCAATTTTTC 1274
 QY 844 TGTTCCTAATTTATGTAGAGGTGGACCGGCTTTTGTCTGAAACGAGTTGGAC 903
 DB 1275 TGTGTCCGAAATCAGAAAAGATGTGGACCGGCTTATCTTTCTCAAAATAGAGTGGTC 1334
 QY 904 AGAATGACGCAATTTTACATTTCTACAAAGTTTGCATGATGATGTTGATGTGAGAGC 963
 DB 1335 GTATGTGTAGANTTTTATGATTTCTATTAATTCAGATTCATGCAAGTATGAGCAACAA 1394
 QY 964 GCAAAAAGACTTGTCTACGCCAAACCAAGTGCAGAGGTGGTATGTTTAAATGGAA 1023
 DB 1395 TTAAAGAAAGATTTATTAAGTTCACATCA-----AATGACGGGGCTAATGTTAAGTGA 1449
 QY 1024 AAACGATCTAGAAATTTCTCCATTTGGACAATTCATACGCAAAACAGTTTGAAGAGT 1083
 DB 1450 CGATGATCTAGAAATTTCTAAATTAAGAAATTTATTCG-AAACCAAGCATGATGAGT 1508
 QY 1084 TACCAAGTTTATTAATGTTTATTTGCGCATATGAGTCTAGTTGGTACAGTCCACTA 1143
 DB 1509 TCGCTCAATTTCTATTAATGTTTAAAGGTGATGATGTTTATGAGAAACGCGCTCCCA 1568
 QY 1144 CAGTTGATGAATTTGAAAAATATATCTCTGTGCAAAAGAGACGATTTGATTTAAACAG 1203
 DB 1569 CAGTTGATGAATTTGAAAAATATATTAATCAACGAGAGACGCGCTTATTTAAACGAG 1628
 QY 1204 GGATTAACAGTCTCTGCGAGGTATAGTGTCTATGATATATCAAGACTTCAGACGATG 1263
 DB 1629 GAATCACTGTGTGTGCGCAAAATATCTGAGAAATATATTAATCTGATTTTGAATGAAATG 1688
 QY 1264 TTCGTTGAGCTTAGCATATCATGATTAATGACATCTGCTAGATTTAAATTTTAT 1323
 DB 1689 TAAAGTTAGATGTTCAATATATCATGAAATGCTCTATTTGTGTGAGATTAATTAATTC 1748
 QY 1324 TAAAGCAGTGAAGTTGTATGTTGAGAAGAGGAAATAGTAAAGTA----- 1372
 DB 1749 TCTTAACGCTAAAGTAGTTTAACTCGGACAGAGACTAAGTAAAGTAAAGTTGAAG 1808
 QY 1373 -----TATGAAGTTGTTGTTGCTCGCTTCTCAAGGGGACATTTGATCTCATTTGATTT 1426
 DB 1809 GAATTAATGAATAATTTGCTGTGTGTTCAAGTGTGTGCACTGTGCACTTGAACCT 1868
 QY 1427 GTTAAACCGTTTGGAGAGAGAAAGAGCTTTTGGGTAACTTTGATTAAGAGAGATGC 1486
 DB 1869 TTTGAAACCCATTTTGGGAAAAAGAAAGATAGTTTGGGTAACTTTGATTAAGAGAGATGC 1928
 QY 1487 AAGAGTCTTTTGAAGATGAAGAAAAATGTATCCATGTTACTTTCCAACAAATGCAATCT 1546
 DB 1929 TAGAGTATTTCTAAGAGAGAGATGTATATCATGCTCTTTCCAACAAACCGTATATG 1988
 QY 1547 CATTAATTTAGTGAATAATCTTTCTTACGTTTCAAAATTTTACGTGATGAGAAACCA 1606
 DB 1989 CAAAAAATTTGTAATAATAATCTATCTAGCTTTTAAAGTCTTAAAGAAAGAACACAGA 2048
 QY 1607 TGTATTATTTATCATCGTGTGGCGCGTGTGCTGCTCTCTTTTAAATGGAAGAACTAT 1666
 DB 2049 TGTATCATATCATCTGTGTGGCGCGTGTGAGACATCTTTTAAATGGAAGTATTT 2108
 QY 1667 TGGAGCAAGACGATTTATTAATGAATTTGATTCAGATTAATTAATTTACATTAACG 1726
 DB 2109 TGGCTGTAGACCGTTTATATAGAGTTTTCGACAGATTAATTAATTAATTTGACAG 2168
 QY 1727 AAAACGATTTATCCGTAAACAGATTTTATTTATGTTCAAGTGGAGAAATGAAGAGT 1786
 DB 2169 AAAATTAAGTATCTGTAAACAGATTAATTTATTTGTTCAAGTGGAGAAATGAAGAAAGT 2228
 QY 1787 ATATCTTAATCTATTAATCTGTGGGAGTATTTTATTAATGATTTTGTAAACAGTAAAGACT 1846

DB 2229 TTATCTTAAGGCAATTAATTTAGAGAAATTTTAAATGATTTTGTGCACAGTAGGACA 2288
 QY 1847 CATGAACAACAGTTAATGATGATTAAGATTAAGATTTATTTGAAAAAATGAGAT 1906
 DB 2289 CATGAACAGAGTTCACCGCTTATTAAGAGTATGATTAAGATTAAGAGGACAGGTCT 2348
 QY 1907 ATACCGAGCAATATTTATTTCAACAGATATTTGATCTATATTTCCAGAAATATTCGAA 1966
 DB 2349 ATTGATCAAGAAATGTTTCAATTTCAACGGTTACTCAGACTTTGAACTCGAATTTGAC 2408
 QY 1967 TATPAAAAATTTCTCAGTTACAAAGAAATGGAACAAATATTAACAAATCAGAGTATG 2026
 DB 2409 TGTCAAAATTTCTCATATGATGATATCACTTTACATGAAGAAAGCTGAGATTTGT 2468
 QY 2027 ATTTGCCAGAGAGCGCCGCTACTTTATGAAATTCATATTCACAAAGAAATTAACATTA 2086
 DB 2469 ATCAACACGGCGGTCCAGCAAGTTTATGAATGCAAGTTTCTAAAGGAAAAAACTAT 2528
 QY 2087 TTGTTCTTAGACAAAAAAGTATGTGAACATGTAATGATCATCAAGTATGAT 2146
 DB 2529 GTGTCTTAGACAAAGACAGTTTGGAGACATGTGAATATCATCAGTGAATTTTGT 2588
 QY 2147 AGAAGATTTTACAGATATAATATTTATTAAGAAAAATATGATGATTTGTTGA 2206
 DB 2589 AAAGATTAATTTGAAATATGATGATTAATTTGAAATATGATGATTAAGAAAT 2648
 QY 2207 AAAATTTGAGTTTCTAAGCAAC---TAACTTTACATCAATATAATTTTGT 2263
 DB 2649 ATTAATTAAGAAAAAATATATCTACTAGTAATGATATCAACAAACATATGATTTTGT 2708
 QY 2264 GAAAGATTAACAAATAGTGAATAATTAATGAGATCAAGAAATGAATTAATAA 2323
 DB 2709 TCTCTTTCAAAATGAACTTTCTTAATCTATTTGAAATATATTTTGTGAGAAAA 2768
 QY 2324 AGATGCA 2330
 DB 2769 AATTTGA 2775

RESULT 7
 ACA64723
 ID ACA64723 standard; DNA; 17276 BP.
 XX
 AC ACA64723;
 XX
 DT 18-JUN-2003 (first entry)
 XX
 DE Streptococcus capular polysaccharide gene.
 XX
 KW Superantigen; ds; gene; SAg; staphylococcal enterotoxin; tumour; cancer;
 KW apoptosis; gene therapy; mammalian cell receptor; cytotoxic;
 KW tumour associated lipid; anergy; T cell; antigen presenting cell; APC;
 KW tumouricidal immunocyte; antitumour.
 XX
 OS Streptococcus sp.
 XX
 PN US2002177551-A1.
 XX
 PD 28-NOV-2002.
 XX
 PF 30-MAY-2001; 2001US-00870759.
 XX
 PR 31-MAY-2000; 2000US-0208128P.
 XX
 PA (TERM/) TERMAN D S.
 XX
 PI Terman DS;
 XX
 DR WPI; 2003-361759/34.
 XX
 PT F-PSDB; AB079115.
 XX
 A mammalian cell receptor, useful in the treatment of cancer by binding

PT to tumor associated lipids where the binding induces anergy or apoptosis
in T cells and antigen presenting cells.

Example 2; Page; 167bp; English.

XX The invention relates to a mammalian cell receptor, useful in the
XX treatment of cancer, which binds to tumour associated lipids and induces
XX anergy or apoptosis in the T cells and antigen presenting cells (APCs).
XX Also included are a mammalian cell useful in the treatment of cancer
XX where the receptor which binds tumour associated lipids and induces
XX cellular inactivation or death is deleted or functionally deactivated,
XX (producing (M1) a tumouricidal immunocyte population in vivo in a mammal
XX by allowing tumour associated lipids to contact immunocytes in which
XX receptors for immunosuppressive fatty acids, ceramides, glycolipids,
XX sphingolipids, glycosphingolipids, phosphosphingolipids, gangliosides,
XX sialylated glycans, lipopeptides and proceoglycolipids are inactivated or
XX deleted), a construct useful in the treatment of cancer comprising a
XX superantigen (SAg) nucleotide inserted into a virus, a mammalian T cell
XX useful in the treatment of cancer (where an adaptor protein which
XX inhibits T cell activation by tumour associated antigens is deleted or
XX functionally deactivated), a composition useful in the treatment of
XX cancer (comprising a lipid raft conjugated to a superantigen), producing
XX (M2) a tumouricidal immunocyte population ex vivo in a mammal (by
XX allowing tumour associated lipids to contact immunocytes, in which
XX receptors for the lipids are inactivated or deleted to produce a
XX tumouricidal immunocyte population, and administering the tumouricidal
XX activated immunocytes to the host), producing (M3) a tumouricidal APC
XX population ex vivo in a mammal (by allowing a tumour associated lipid to
XX contact APCs, in which receptors for the tumour associated lipids are
XX inactivated or deleted to produce a tumouricidally activated population,
XX and administering APCs to the host), producing a tumouricidal T cell
XX population ex vivo in a mammal) by allowing a tumour associated lipids to
XX contact T cells, in which adaptor proteins, which inhibit T cell
XX activation by tumour associated antigens, are deleted or functionally
XX deactivated to produce a tumouricidal population of T cells, and
XX administering the tumouricidally activated T cells to the host, or
XX allowing a superantigen-lipid raft to contact T cells ex vivo, and
XX administering the tumouricidally activated T cells to the host), treating
XX (M5) cancer in a mammal (by administering a lipid binding molecule which
XX binds immunosuppressive tumour associated lipids in vivo), producing (M6)
XX a tumouricidal T cell population in vivo in a mammal (by allowing a
XX tumour associated antigen to contact immunocytes in which adaptor
XX proteins which inhibit T cell activation by tumour associated antigens
XX are deleted or functionally deactivated) and producing (M7) a
XX tumouricidal T cell population ex vivo in a mammal comprising allowing a
XX superantigen-lipid raft conjugate to contact immunocytes in vivo. The
XX receptors, methods and compositions are useful for treating cancers and
XX tumours. Bacterial superantigens are co-administered or administered as
XX fusion constructs with anti-tumour proteins or motifs. The present
XX sequence encodes an anti-tumour protein which is co-administered with or
XX incorporated into a fusion construct with a superantigen. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format from the US patent
XX office website at "seqdata.uspto.gov/sequence.html?docID=20020177551"

XX Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;

XX Query Match 10.4%; Score 726.8; DB 8; Length 17276;

XX Best Local Similarity 59.2%; Pred. No. 1e-88;

XX Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

QY 16 TGGCATTTATGATGATGAGTCGACATTTCTGCAATCTTAACAAGTCATATACCA 75
DB 4765 TACGAGTATTCACAAAGCTGTGATTTATTTCTGCAAGTTGACATTAATTA 4824
QY 76 ATGCTGATTTAAATCG---TTCTGCAATTTTATCATATGATGCTCATTTATTTGCAT 132
DB 4825 CTCCTCAATTTTAAAGCAATTAAGATTTATTTGTTCTATTTGATACATTAATATGTTT 4884
QY 133 TTTTATATCTCGTATGCCAGTTGAATTTGATGATAGGATTAATCTGATAGAGTTTGAA 192
DB 4885 TTATATCTTTCGATTTTACAGAGACTTTTGAGAGCTGCTATCTTGAAAGAGTTTAAA 4944

QY 193 AAACATTTAATGATATGATTAATATTTGCAATTTTCTTACGCGACATATCTTTGCTG 252
DB 4945 TGGTATGAAAATACAGCTTTTACTATATTTTCAATACAGTTCAATATTTTATTTTAA 5004
QY 253 AGAATTAATTTGGACCTTTCAAGAGCTGTGCGGTATTTTCAATTAATTAACCTTGTT 312
DB 5005 AAAACTCTTTTACACGACACGACCTTCTTTTACTTTTATTTGCTATGAAATTCGATTT 5064
QY 313 TGGTATTAATTTTAAACGTAATTTATTTAGCAGTTTAAAGATAGCTTTCTATTTTGAACA 372
DB 5065 TATTAATCTATTTGAATTCATTTTAAATATTTATCGAAAATATCTTAACGTAAGTTT 5124
QY 373 TCTATCAAAAAAAGACATTTCTAATTAACAAGCGCTGAACATGGGAAAATATGCAAGTTT 432
DB 5125 CACGAGATACCAAAAGTTTGTGTAAGAAATTAAGAAATCTTTATCAAAAAGACCTTTA 5184
QY 433 TATTTGAATTCATTAACAAATTCAAAATAATCTGTGATCTGTAGTTTATGAGTACAG 492
DB 5185 GGAATTAAT--ACACCATTAATTTATTCGCTGTCTGATCTTGACATCTCTGAAAAGG 5241
QY 493 AAATGATTAATAATTTATTCATTAACGCTCTATTAATTTCTGTGGAAGAGCTATAGAGT 552
DB 5242 ATGTATATGATTTGAAACATTAACCTGTTAAGGATTAATTAACAAAGATGCTTACTTCAG 5301
QY 553 TTTCAACAAAGGAAGTGTGACACGCTTTATTAATCTAACCAAGTATTTAGACG 612
DB 5302 AGTTAACTGCTTAACTGTGATACAGCTTTTATTAACATACCAATGAAATATTTTGTA 5361
QY 613 TTAACGA-----TTGCTTACAGATTTTGAAGTTTGAAGTTATGATGAACGTTG 663
DB 5362 AATTAACCAATACAGATTAATTAATTAATGACATTTGAAGCAATGGAATGATTCATGTTA 5421
QY 664 ATATTAAATTCATTCGATTTTCTGCTGTTGAAAACAAAATAATCAACTGCTAGGTACC 723
DB 5422 ATGTAGAGCACTTAGCTTTGATTAATATGAGAAAGAAATCCAAACTTTTGAAGAT 5481
QY 724 ATAGCATTTAACTTTTTCACAAATTTTATTAAGCTCTGATATCAATGATGAAGACG 783
DB 5482 ATAGCTTATTTACATATTTCTATGAAATTTCTAATAATATGATCACTTAATGACAAACGAT 5541
QY 784 TTTTGGATTTACTCGGAGCGGATAGCTGGGTAAATTTATTTGGTATTTGTTCTATTTTGT 843
DB 5542 TTTTGGATTTACCGGCTGTATTAATAGTTTGTCTATATGTCATTTGGCAATTTTTC 5601
QY 844 TAGTTCCATTTATTCGTATGAGATGTGACCGGCTATTTTGTCTGAAAACGATTTGAC 903
DB 5602 TAGTTCCGAATTCAGAAAAGATGTGACCGGCTATTTTCTCAAAATAGATGTGTC 5661
QY 904 AGAATGACGATATTTTACATTTTACAGATTTTCAATGATGATATGATGTCGAGAGAC 963
DB 5662 GTAAATGTATGAGATTTTATGATTTCTATTAATTTCAATGATGCAATGCAATGCAACAA 5721
QY 964 GCAAAAAGACTGTCACACCAAAAACAGATGCAAGGTGGGTATTTTAAATGGGAA 1023
DB 5722 TTAAAGAAAGATTTATGATTTTCAATCAATCAATGACAGG-----CTAATGTTTAAGTTA 5776
QY 1024 AAACGATCTTACATTTACTCAATTTGACATTTTCAATGCAAAAACAGTTTAGACGAGT 1083
DB 5777 AGATGATCTTCAATTTACTTAATTAATAGAAATTTATTCG-AAAAACAAGCATGATGAGT 5835
QY 1084 TACACAGTTTATTAATGTTTATTTATTTAGCGATATGATGCTAGTTGTCACGTCACCTTA 1143
DB 5836 TGCCTCAATTCATATATGTTTAAAGCGATGTGAGTTTACGAGGAACCGCCTTCCA 5895
QY 1144 CAGTTGATGATTTGAAAAATATATCTCTGTGTAAGAGAGATGAGTTTAAACGAG 1203
DB 5896 CAGTTGATGATTTGAAAAATATATTAATTAATCAACGAGAGCGCTTATGTTTAAAGCGAG 5955
QY 1204 GATTTACAGTCTCTGCGAGGTTAGTGTGCTGATATATCAACAGCTTCGACGAGCTAG 1263
DB 5956 GAATCACTGTGTTTGTGCAAAATATCTGTGTAAGAAATATATTAATGATTTTGAAGAATCG 6015
QY 1264 TTGCTTTGCACTTAGCATATCATTTGATTAATTTGAATCTGCTGATATTTAAATTTTAT 1323


```

Db      6016 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTTGGTCAGATATTAAGATTATTC
      1324 TAAAGACAGTAAGTTGATTTGTGAGAGAGAGTAAGTAAAGTA-----1372
      6076 TCCTAACACTAAAGGAGTAGTTTACTCGGACAGAGACTAAGTAAAGGTTTGAAG 6135
      1373 -----TATGAAAGTTTGGTTCGGTCTTCAAGGGGACATTGTCATCTGATTTT 1426
      6136 GATATATATGAAATTTGCTGTGGTTGGTTCAAGTGGTGCATCTAGACACACTTGAACCT 6195
      1427 GTTAAACCGTTTGGAGAGAGAACGTTTTGGGTAACATTTGATTAAGAGATGC 1486
      6196 TTTGAAACCATTTGGGAAAAAGAGATAGTTTGGGTAACTTTGATTAAGAGATGC 6255
      1487 AAGAACTCTTTGAAGATGAAAAAATGATCCATGTTACTTTCCAAATGCGCAATCT 1546
      6256 TAGAGATATTTCAAGAGAGAGATGATATATCATTTGCTTCCAAACCGTATATGT 6315
      1547 CATTAATTTAGTAAAAATCTTTCTAGCTTCAAAATTTTACGATGAGAAACGACA 1606
      6316 CAAAACTTGTTAAAAAATATCTATTTAGCTTTTAAAGTCTTGAAGAAAGACACGA 6375
      1607 TGTATATATTTCAATCTGTGCGCGCGTGTCTGCCCTTCTTTACATCGAAACTAT 1666
      6376 TGTATCATATCATCTGTGCGCGCTGAGACATCACTTTCTTTATATGTTAATTT 6435
      1667 TGGAGCAAGACATTTATTTGAATTTGATTCAGATTAATTAATCTACATTAATCG 1726
      6436 TGGTTGTAGACCGGTTTATATAGAGTTTTCGACAGATAGATTAACCACTTTGACAG 6495
      1727 AAAACTAGTTTATCCGTAACAGATTTTATTTGTTCACTGGGAAAGAAATGAAGAAGT 1786
      6496 AAAATTAGTATCTGTAACAGATTAATTTATTTGTCAGTGGGAAAGAAATGAAGAAGT 6555
      1787 ATATCCTAATCTATTAATCTGAGGAGTATTTTATGATTTTGTAAACATGAGAACT 1846
      6556 TTAATCTTAAGCAATTAATTTAGAGAGAAATTTTATATGATTTTGTACAGTGGGACA 6615
      1847 CATGAACAAGTTTATGATGATGATTAAGAGATGATTTTATGAAAAAATGAGAGT 1906
      6616 CATGAACAAGTTTATGATGATGATTAAGAGATGATTTTATGAAAAAATGAGAGT 6675
      1907 ATTAACGAGAAATATTTATTTCAACAGATTTCTGACATATTTCCAGAAATTTGCCAG 1966
      6676 ATTAATCAAGAAAGTTTCAATTTCAACGAGTTTCTGACATTTGCAACCTCGAAATTTG 6735
      1967 TATAAAAAATTTCTCAGTTACAAAGAAATGAAACAATATATTAACAATCAGAAGT 2026
      6736 TGTCAAAATTTCTCATATGATGATGATGATCTTACATGAAAGAGCTGAGATTTGT 6795
      2027 ATTTCGACGAGAGCGCCGCTACTTTTATGAAATTCATTTCAAGAGAAAAAACAATTA 2086
      6796 ATCAACATATGGCGCCGACGAGCGTTATGTCAGTTATTTCTTAGGAAATTAACAGTT 6855
      2087 TTGTTTCTTACAGAAAAAAGTATGTTGAACATGTAATGATCATCAAGTAGGTTTGA 2146
      6856 GTTGTCTTACAGAGAAAGAGTGTGTAACATATGATCATCAAAATTAACAATTTTGA 6915
      2147 AGAAGATTTTACAGATATATATTTATTTATGAAAAATATAGTATTTGTTGAA 2206
      6916 AAAAAAATTTGCCACCTGATTCCTTGGCTTGGATTTGAAGATTAAGTGAATTTGGGAA 6975
      2207 AAAATTTATGAAGTTTCTAAGCAAACTAATTTACATCAAAATTAATTTTGTGGA 2266
      6976 GCGTTAAAGAAATTAAGCTACAGAAAAATATCAGGAAATATATGATGTTTGTGCAT 7035
      2267 AGATTTAAACAAATAGTTGAAAAAT 2292
      7036 AAATTAGAAAAAATTAATAGTGAAT 7061

```

```

ADFA3363
ID      ADF43363 standard; DNA; 17276 BP.
XX
AC      ADF43363;
XX
DT      12-FEB-2004 (first entry)
XX
DE      Streptococcus capsular polysaccharide DNA seq id 83.
XX
KW      receptor; lipid-based tumour associated antigen; cytostatic;
KW      antimicrobial; gene therapy; neoplastic disease; tumour; cancer;
KW      infectious disease; capsular polysaccharide; ds.
XX
OS      Streptococcus.
XX
PN      US2003157113-A1.
XX
PD      21-AUG-2003.
XX
PF      28-DEC-2000; 2000US-00751708.
XX
PR      28-DEC-1999; 99US-0173371P.
XX
(TERM/) Terman D S.
XX
Terman DS;
XX
WPI; 2003-787326/74.
XX
P-P-SDB; ADF43364.
XX
PT      New receptor in a mammalian cell that inhibits regular activation by
PT      receptors specific for lipid-based tumor associated antigens, useful for
PT      treating a neoplastic disease or tumor, and infectious diseases.
XX
Example 3; SEQ ID NO 83; 151bp; English.
XX
PS
CC      The invention describes a receptor in a mammalian cell that inhibits
CC      regular activation by receptors specific for lipid-based tumour
CC      associated antigen. The receptor has cytostatic and antimicrobial
CC      properties and is suitable for use in gene therapy. The receptors,
CC      methods and compositions are useful for treating a neoplastic disease or
CC      tumor (cancer), and infectious diseases. This sequence represents a
CC      streptococcal capsular polysaccharide polynucleotide, a cell surface
CC      moiety, the DNA of which can be transfected into a cell with superantigen
CC      DNA to generate antitumour immunity.
XX
SQ      Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;
XX
Query Match      10.4%; Score 726.8; DB 10; Length 17276;
Best Local Similarity 59.2%; Pred. No. 1e-88;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;
XX
      16 TGGCATTTATTTGATGATGATGAGCAATTTCTGCAATCTTACAGATCATATACAA 75
      4765 TAGCATGATTTCAACAGTTGTAGTTATTTTCTGCAAGTTGACATTAACATTAATTA 4824
      76 ATGCTGATTTAAATCG---TTCTGAAATTTTATCATATGATGCTTCATATTTTGCAT 132
      4825 CTCCTCAATTTTAAACAAATTAAGATTTATGTTTGTCTATTTGATATATATATTTGTT 4884
      133 TTTTATATCTGATGATGCAAGTTGAATTTGAGTATGAGGTATCTGATGAGTTGAA 192
      4885 TTTATCTTCTGATTTTACAGAGACTTTTGAGTGTGGCTATCTTGAAGAGTTTAA 4944
      193 AAAATTTAATGATGATATATTTTGAATTTTCTTACGAGATCATTTTGTGG 252
      4945 TGTATTTGAAATACAGCTTTTACTATATTTTCAATATCAAGTTCATATTTTATTTTGA 5004
      253 AGAATAATTTTGGACTTCAAGACGTCGTCCTGATATTTCAATTAATTAATTTGCTT 312
      5005 AAAACTTTTACACAGACGACCTTTCTTTTATCTTTATTTGCTATGAATTCGATTT 5064
      313 TGGTATACCTATTTAAGTATATTTAAGCAGTTTAAGATAGCTTTCTATTTTGCACA 372

```

Db 5065 TATATATCTATGTAATTCATTTTAAATAATATTCGAAAATATTCCTAACGTAAGTTT 5124
Qy 373 TCTATCAAAAAAAGCAGATTCTAATTTACACGGCTGGAAGATGGAAAAATATGCAAGTT 432
Db 5125 CACGAGATACCAAGATTGTTTGTATACGAATAGAGATTTCTTATCAAAAAATGACCTTGA 5184
Qy 433 TATTTGAATCATATAACAAATTCGAAAAATCTTGTTCATTTGGTATGGATTTTGGTACG 492
Db 5185 GGAATTAAT--ACGACCTAATATATGCTGTCTGTATCTTGGACCTCTCGAAAAAG 5241
Qy 493 AAATAGATTAATAATTAATTTATCATTAACGCTCTATATATCTGTGGAAGAGCTATAGACT 552
Db 5242 ATTGTATGATTTGAAACATACTCGTTAAGGTAATAAACAAGATGCTCTTACTTCAG 5301
Qy 553 TTTCAACAAGGAGAGTGTGCACACGCTTTTATTAATCTACCAAGTGGTTTATAGAC 612
Db 5302 AGTTAACTGCTTAACGTGTGATCAAGCTTTTATTAACATACCATTGAAATTAATTTGGTA 5361
Qy 613 TTAAGCA-----TTCGTTTGAGATTTTGAAGTTGTAGGATTTGATGAGCGTTG 663
Db 5362 AATPACCAATACAGATATTTATTAATGACATTTGAAGCAATGGAGATGATTCATATGTTA 5421
Qy 664 ATATTAAATTCATTCGCTTTTACTGCGTGAACCAAAAAATCCAACTGCTAGTGACC 723
Db 5422 ATGTAGAGCACTTAGCTTTGATATATAGAGAAAAAGCAATCCAACTTTTGAAGAT 5481
Qy 724 ATAGCATGTAACTTTTCCAAATTTTATTAAGCTAGTCAATATCATAGATGAAGACG 783
Db 5482 ATAGTGTTTATCATATTTCTATGCAAAATTCATTAATATATAGTCCCTTAATAGCAAAACGAT 5541
Qy 784 TTTTGGATATACCTCGAGCGGTAGTGGGTTAATTTATTTGGTATAGTTTCTAATTTGT 843
Db 5542 TTTTGGATATACGGGTGCTATTTATAGGTTTGTCTATGTGCAATTTGGCAATTTTTC 5601
Qy 844 TAGTTCCAAATTAATTCGTAGATGTGACCGGCTAATTTTGTCTGAAAAAGATTGAC 903
Db 5602 TAGTTCCGAAATACAAAAAGATGTGACCGGCTAATTTTCTCAAAATAGAGTGTGTC 5661
Qy 904 AGATGACGCAATTTTACATTTCTACAGATTTGATGATGTATGTGATGCTGAGAGAC 963
Db 5662 GTATATGTAGATTTTATAGATTCATTAATTCAGATCAATGAGATGAGAGCAAA 5721
Qy 964 GCAAAAAGACTGCTCAGCCAAACCAAGATGCAAGGTTGGTATGTTTAAATGGAA 1023
Db 5722 TTAAAGAAAGTTATTAATGTTCAACATCAAAATGACAGG-----CTAATGTTAAGTTAA 5776
Qy 1024 AAACGATCCTAGATTAATCTCAATTTGACATTTCAATACGCAAAAACAGTTTACAGAGT 1083
Db 5777 AGATGATCTAGATTAATCTAAATAGGAAATTTAATTCG-AAAAACAAGATGATGAGT 5835
Qy 1084 TACCAAGATTTTAATGTTTAAATTTGCGATATGAGTCTAGTTGGTACAGCTTCACCTTA 1143
Db 5836 TGCTCAATCTAATATGTTTAAAGGCGATATGAGTTTGAAGGAAACGCGCCCTCCA 5895
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCGTCAAAAAGAGACGATTTGATTTAAACAG 1203
Db 5896 CAGTTGATGAATTTGAAAAATATATCAACGCAAGGACGCGCTTAGTTTAAAGCCAG 5955
Qy 1204 GATTTACAGGCTCTCGGACGTTAGTGTGATGTAATATACAGACTTCGACGACGCTAG 1263
Db 5956 GAATCACTGTTTGTGCAAAATATCTGTAGAAATATATTACTGATTTTGGATGAATCG 6015
Qy 1264 TTGCGTTGACTTAGCATACATTAATTAATTTGACTATCTGTGCTAGATTTAAATTTAT 1323
Db 6016 TTAAGTTAGATTTCAATATATCAATATGATGATGATTTGCTAGATTTAAGATTAATTC 6075
Qy 1324 TAAAGACATGAAGTGTATGTTGAGAGAGGAAATGAAGTA----- 1372
Db 6076 TCTTAACACTAAAGTATGTTTACTCGGACAGGAGCTAAAGTAAAGGTTTGAAG 6135
Qy 1373 -----TATGAAAGTTGTTGTGCGGTCTTCAGGCGGACATTTGACTCATCTGTATTT 1426

Db 6136 GAATTAATTAAGAAATTTGCTGTGTTGTTCAAGTGTGTATCTAGACACTTGAACCT 6195
Qy 1427 GTTAAACCGTTTGTGAAGAAGAAAGCTTTTGGGTAAACATTTGATTAAGAGATGC 1486
Db 6196 TTTGAACCCATTTTGGAAAAAAGATAGGTTTGGGTAACTTTGATTAAGAAGATGC 6255
Qy 1487 AAGAAGCTTTTGAAGATGAAGAAATGTATTCATGTTACTTCCAAAGATCCGATGCT 1546
Db 6256 TAGAGTATTTCTAAGAGAAAGATGTATATCATTTGCTTTTCCAAACACCGTATGT 6315
Qy 1547 CATTAATTTAGAAAAATCTTTCTAGCTTTCAAAATTTTACGTGATGAGAAACAGA 1606
Db 6316 CAAAACTTGTGAATAATCTATTTCTAGCTTTTAAAGTCCCTTGAAGAAAGACCGAG 6375
Qy 1607 TGTATTAATTTATCTGTGTGCGCCGTGCTGTCCCTTTCTTTTATCATCGAAAACTAT 1666
Db 6376 TGTATCATATCATCTGTGTGCGCGGTGAGCAGTACATTTCTTTTATTTGGTAAATTT 6435
Qy 1667 TGAACCAAGACGATTTTATTTGAAGTATTTGATGAGTTAATTAATCTACCTTAACCTG 1726
Db 6436 TGGTTGTAAAGCCGTTTATATAGAGTTTTCGACAGATAGATTAACCAACTTTGACAG 6495
Qy 1727 AAACCTAGTTATCCGTAACAGATATTTTATTTGTCAGTGGAGAAATGAAGAGT 1786
Db 6496 AAAATTAAGTATCTGTACAGATTAATTTATTTTATGATTTTGTCACTGTGGGACA 6555
Qy 1787 ATATCTTAATCTATTAACCTGGAGATTTTATATGATTTTGTATCACTAGAACT 1846
Db 6556 TTATCTTAAGCAATTTAATTTAGAGGAATTTTATGATTTTGTCACTGTGGGACA 6615
Qy 1847 CATGAACAAGTTTATTCGATTTGATTAAGAGATTTGATTTTGAAGAAATGAAGT 1906
Db 6616 CATGAACAAGCTTCAACGCTTTATTAAGAGATTTGATAGTATTAAGGACAGGTGCT 6675
Qy 1907 ATPACGAGAAATTTTATTCAAAACAGATATTTCTGATATATTCAGAAATTTGCAAG 1966
Db 6676 ATTTGATCAAGAGATTTTATTAACAGGTTTCTGACCTTGAACCTCAGATTTGTCAG 6735
Qy 1967 TATTAATAATTTCTCAGTTACAAAGAAATGAGAAATATATTAACAAATCAGAGTGT 2026
Db 6736 TGGTCAAAATTTCTCAGATATGATGATGATGATGATGATGATGATGATGATGAT 6795
Qy 2027 ATTTGCGACGAGGCGCCGCTTCTTTATGAATCATTAATCCAAAGGAAATTAACATTA 2086
Db 6796 ATCACACATGCGGCGCCACGCACTTTATGTGATGATTTCTTTAGGAAATTTACAGTT 6855
Qy 2087 TTGTTCCCTAAGCAAAAAAGTATGTGAAACATGTAATGATCATCACTAGGTTGTA 2146
Db 6856 GTTGTCTTAGAGAAAGAGCTTTGTGAAACATATCATATGATCATTAATCAATTTTGA 6915
Qy 2147 AGAAGAAATTTTCAAGATATATATTTTATTAAGAAATATAGATATTTGTTGA 2206
Db 6916 AAAAAAATTTGCCCACTGTATCCCTTGGCTGGAATTTGAAGATGATGATGATGATGAT 6975
Qy 2207 AAAATTAATGAAGTTTCTAAGCAATTAATTTTATCAATCAATTAATTTTGTGAA 2266
Db 6976 GCGTTGAAAAAGAAATATAGCTACAGAAAAATATCAGGAAATATATATATGTTGTAT 7035
Qy 2267 AGATTAAACAAATAGTTAAAAAT 2292
Db 7036 AAATTGAAAAAATTAATAGGTAAAT 7061

RESULT 9

AEA03034 standard, DNA, 17276 BP.

AEA03034;

28-JUL-2005 (first entry)

Staphylococcal hemolysin nucleotide sequence SEQ ID NO:60.

KW tumor; neoplasm; gene therapy; immunotherapy; cytostatic; hemolysin;
 KW gene; ds.
 XX
 OS *Staphylococcus* sp.
 XX
 PN US2005112141-A1.
 XX
 PD 26-MAY-2005.
 XX
 PF 08-SEP-2004; 2004US-00937758.
 XX
 PR 30-AUG-2000; 2000US-00650884.
 XX
 PA (TERM/) TERMAN D. S.
 XX
 PI Terman DS;
 DR WPI; 2005-394926/40.
 DR P-PSDB; AEA03035.
 XX
 PT New composition for treating a tumor or neoplastic disease in a subject
 PT comprises conjugates comprising superantigen polypeptides or nucleic
 PT acids with other molecules that produce a tumoricidal response.
 XX
 PS Example 3; SEQ ID NO 60; 125bp; English.
 XX
 CC The invention relates to a composition for treating a tumor or neoplastic
 CC disease in a subject. Also described: (1) a mammalian cell comprising an
 CC exogenous nucleic acid encoding a superantigen expressed in the cell,
 CC which cell also produces or expresses all alpha-anomers of
 CC monoglycosylceramide or diglycosylceramide, where expression of the
 CC superantigen and the mono- or diglycosylceramide is capable of eliciting
 CC an antitumor immune response in a mammal into which the cell is
 CC introduced; (2) treating a tumor or neoplastic disease in a subject; (3)
 CC preparing a population of immunotherapeutic T or natural killer T (NKT)
 CC cells useful to treat a tumor or neoplastic disease in a subject; (4) an
 CC apoptotic cell preparation or lysate useful for treating a tumor or
 CC neoplastic disease in a subject, comprising a cell population that has
 CC been transfected with naked DNA encoding a superantigen, and treated to
 CC undergo apoptosis or lysis; and (5) a cell that has ingested or been
 CC transfected with the above apoptotic preparation or lysate, thus,
 CC rendering the cell effective in presenting material expressed from
 CC transfecting nucleic acid or material ingested to the immune system of a
 CC mammal to elicit an anti-tumor immune response. The composition and
 CC methods are useful for treating tumors or neoplastic diseases. The
 CC present sequence represents a hemolysin nucleotide sequence, which is
 CC used in an example from the present invention. Note - The sequence data
 CC for this patent is not represented in the printed specification, but was
 CC obtained in electronic format directly from the USPTO web site.
 XX
 SQ Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;
 Query Match 10.4%; Score 726.8; DB 14; Length 17276;
 Best Local Similarity 59.2%; Pred. No. 1e-88;
 Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

DB 5005 AAAAATCTTTTACAGCAGCAGCTTCTTTTCTTTTATCTTATGATGATTCGATTT 5064
 QY 313 TGGTATACCTATTTAAGTATATTTAAGAGTTAAGATAGCTTCTATTTTGCACAA 372
 DB 5065 TATTATATCTTATGAAATTCATTTTAAATATATATGAAATATTTCTTACGCTAAGTTT 5124
 QY 373 TCTATCAAAAAAGACGATTCATTTTCAACGCGCTGACGATGGAATAATATGCAAGTTT 432
 DB 5125 CAGAGATACCAAGTTGTTTGTATACGAATAAGATTTCTTATCAAAAAATGACCTTTA 5184
 QY 433 TATTGATGACATTAACAAATTCAAAAAATCTTGTGATGTGATGTTTATAGTACAG 492
 DB 5185 GGAATTAAT--ACGACATTAATTAATTCCTGCTGTATCTTGAATCTCCTGAAAAAG 5241
 QY 493 AAATGATTAATTAATTTATTCATTCACCTGCTATTTATCTGCGAAGAAAGCTATAGT 552
 DB 5242 ATTGTATGATTTGAAACATTAATCTGTTAAGATTAATTAACAAAGATCTTTACTTCAG 5301
 QY 553 TTTCAACAAAGGAAGTGTGACACGCTTTTATTAATCTTACCAAGTATTTTACAG 612
 DB 5302 AGTTAACCTGCTTAACTGTTGATGACAGCTTTTATTAATCAATACCATTAATTTGCTA 5361
 QY 613 TAAAGCA-----TTGTTTCAAGATTTTGTAGTTAGTATGATGATGATGATG 663
 DB 5362 AATACCAATACAAAGATATTAATTAATGACATTAAGCAATGGAGTATGTCATGTTA 5421
 QY 664 ATATTAATTCATTCGTTTGTACGCTGTGAAAAAACAATAATCACTGATGATGAC 723
 DB 5422 ATGTAGAGGCACTTAGCTTGAATTAATTAATGAGAAAAAGCAATCTTTGAAAGAT 5481
 QY 724 ATGCAATGTAATCTTTTCCAAATTTTATTAAGCCTATGATGATGATGATGATGATG 783
 DB 5482 ATATGTTATTAATCAATATCTATGAAATCTATTAATTAATTAATTAATTAATTAAT 5541
 QY 784 TTTTGTATATCTCGAGCGGTAGTGGGTTTATTTTGTATGATGATGATGATGATGATG 843
 DB 5542 TTTTGTATATCAGGGGTGATTAATTAATGATGATGATGATGATGATGATGATGATG 5601
 QY 844 TAGTTCATTAATTCGTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 903
 DB 5602 TAGTTCGCAAAATCAGAAAAAGATGATGATGATGATGATGATGATGATGATGATG 5661
 QY 904 AGATGAGCAGATTAATTAATCTCAAGATGATGATGATGATGATGATGATGATGATG 963
 DB 5662 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5721
 QY 964 GCAAAAAAGACTTGTCTCAGCCAAACAGATGCAAGGATGATGATGATGATGATGATG 1023
 DB 5722 TTAAGAAAGATTTATTAATGATGATGATGATGATGATGATGATGATGATGATG 5776
 QY 1024 AAAGCATCTAGAAATTAATCTCAATGATGATGATGATGATGATGATGATGATGATG 1083
 DB 5777 AGATGATCTAGAAATTAATTAATGATGATGATGATGATGATGATGATGATGATG 5835
 QY 1084 TACCAAGTTTATTAATGTTTATTTGCGATGATGATGATGATGATGATGATGATGATG 1143
 DB 5836 TGCTCAATTCATTAATGTTTATTTAAAGGATGATGATGATGATGATGATGATGATG 5895
 QY 1144 CAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1203
 DB 5896 CAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5955
 QY 1204 GATTAACAGTCTCTGCGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1263
 DB 5956 GATTAACAGTCTCTGCGAGGATGATGATGATGATGATGATGATGATGATGATGATG 6015
 QY 1264 TTAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1323
 DB 6016 TAAAGTTAGTGTCAATATATCAATGAAGATGATGATGATGATGATGATGATGATG 6075
 QY 1324 TAAAGCAGTGAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1372

DE Streptococcus agalactiae DNA sequence, SEQ ID 135.
 XX Antibacterial; vaccine; bacterial infection; ds.
 KM Streptococcus agalactiae.
 XX MO200292818-A2.
 XX 21-NOV-2002.
 XX 26-APR-2002; 2002MO-1B003059.
 XX 26-APR-2001; 2001FR-00005642.
 XX (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX Glaefer P, Rusniok C, Chevalier F, Frangeul L, Lalloué L,
 XX Zouine M, Couve E, Buchrieser C, Poyart C, Txiu-Cuot P, Kunst F,
 XX WPI; 2004-101891/11.
 DR Genomic nucleotide sequences encoding polypeptides of Streptococcus
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
 PT and identification of therapeutic targets.
 XX Claim 1; SEQ ID NO 135; 439pb; French.
 XX The present invention relates to novel Streptococcus agalactiae
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
 CC synthesis of amino acids, cell membranes, intermediate (central)
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
 CC regulatory functions, replication, transcription, translation, protein
 CC transport, adaptation to atypical conditions, sensitivity to medicines
 CC and/or analogues, functions related to transposons, biosynthesis of
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and
 CC cellular machinery. (I) are useful for the detection and/or amplification
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
 CC useful for treatment of a bacterial S. agalactiae infection. The complete
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
 CC present patent is an equivalent for the basic patent FR2824074A1, which
 CC contains only 2344 sequences.
 XX
 XX Sequence 95596 BP; 31762 A; 16282 C; 17030 G; 30522 T; 0 U; 0 Other;
 SO
 Query Match 10.4%; Score 725.2; DB 13; Length 95596;
 Best Local Similarity 59.2%; Pred. No. 1.6e-88;
 Matches 1365; Conservative 0; Mismatches 903; Indels 38; Gaps 6;

DB 3873 TATTGTATCTATTTGTAATTCATTTTAAATATATATCGAAATATTTCTTACGTAAGTTT 3932
 QY 373 TCTATCAAAAAAAGACGATTCCTATTTACACGCGTGAACGATGGAAAAATATGCAAGTT 432
 DB 3993 CACGAGATACCAAGATGTTTGGATACGAATAGAGATTTCTTATCAAAAAATGACCTTTA 3992
 QY 433 TATTTGAATCACTAATTAACAAATTCAAAAATCTTGTGCAATGCTAGTTTATGCTACG 492
 DB 3993 GGAATTAAT---ACAGCATATATATATCGCTGTCTGATCTTGGACATCTCTGAAAAAG 4049
 QY 493 AATAGATTAATTAATTTATATCATTCACGCTATATTTCTGTGGAAGAGCTATAGAGT 552
 DB 4050 ATTGTTATGATTTGGAACATTAACCTGTTAAGATTAATTAACAAAGATGCTTACTTACG 4109
 QY 553 TTTCAACAAGGAGGATGTCGACCAAGCTTTATTAATCTTACCAAGTATTTTATAGAG 612
 DB 4110 AGTTAACCTGCTTAACCTGTGATTCAGCTTTTATTAACATACCAATGATTTATTTGAT 4169
 QY 613 TAAAGCA-----TTCGTTGAGATTTTGAAGTTTATGATTTGATGTAAGCTTG 663
 DB 4170 AATACCAATATCAAGATATATTAATGACATTTGAAGCAATGGAGTGTCTCAATGTAT 4229
 QY 664 ATATTATTCATTCGCTTTTACTGCGTTGAATAACAAAAATCCAACTGCTAGAGTAC 723
 DB 4230 ATGTAGAGGCACTTACCTTTGATTAATATAGGAAAAAGCAATCCAACTTTGAAGAT 4289
 QY 724 ATAGCATGTAACTTTTCCACAAATTTTATTAAGGCTAGCATATCATATGATGAACGAC 783
 DB 4290 ATATGCTTTATTCATATTTTATGAATTTTATTAATATATGTCATACCTTATAGCAAAACGAT 4349
 QY 784 TTTTGATATCTCGAGGAGGAGTACGCGGTATTAATTTTGTGATATGTTCTATTTTGT 843
 DB 4350 TTTTGATATCAAGGCTGTATATATAGTTTCTCATATGTGTCATTTGCGCAATTTTTC 4409
 QY 844 TAGTTCCAAATTTATCTATAGATGTGTGACCGGCTATTTTGTCTGGAACGAGTTGAC 903
 DB 4410 TAGTTCCAAATTTATCAAAAAAGATGTGACCGGCTATTTTCTCAAAATATGATGATG 4469
 QY 904 AGAATGAGGCAATTTTATCACTTCAATCAAGTTTCGATCGATGTATGATGTCGAGAGC 963
 DB 4470 GTATGTATGAGATTTTATGATCTATTAATTTCAATCAATGTCATGATGATGATGATG 4529
 QY 4530 TTAAGAAAGATTTATATGTTTCAATCAATGACAGG-----CTAATGTTTAAAGTTAG 4584
 QY 1024 AATGATCTTGAATTAATCTCAATTTGACATTTGACATTCATACGCAAAACAAGTTTGAAGAT 1083
 DB 4585 AGATGATCTTGAATTAATCTCAATTTGACATTTGACATTTGACATTTGACATTTGACAT 4643
 QY 1084 TACCAAGTTTATTAATGTTTATTAATGCGGATATGAGTCTAGTTGACATGTCACCTTA 1143
 DB 4644 TGCTCAATCTAATATATATATTAATTAAGGATATGATGATTTAGTATGAAACACGCTTCCA 4703
 QY 1144 CAGTTGATGATTTGAAAAATATATCTCTGCTCAAAAGACGATTTGAACGAG 1203
 DB 4704 CAGTTGATGATTTGAAAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4763
 QY 1204 GATTTACAGTCTCTGCGAGGTTAGTGTGTATGATATATCAACAGCTTGAAGAGCTAG 1263
 DB 4764 GAATCACTGTTGTGCAAAATATCTGTGATATATATCACTGATTTTGAAGAAATGCG 4823
 QY 1264 TTTGGTTGACCTTGAATCATTTGATTAATTTGATGATCTGCTGAGATTTAATTAATTTAT 1323
 DB 4824 TAAAGTTAGTTTCAATATATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 4883
 QY 1324 TAAAGCAGTGAAGTTGATTTGTTGAGAGAGGAAATGATTAAGTA----- 1372
 DB 4884 TCTTACACATTAAGGATGCTTACTTGGACAGGCTGATTAAGTAAGTATGATTTGAAG 4943
 QY 1373 -----TATGAAGTTTGTGTGCTTCTTCAAGGAGGACATTTGATCACTTGTATTT 1426

Db 4944 GAATATATGAAAAATTTGCTGGTGTGCTCAAGTGTGTGCTATGACACATTTGAACCT 5003
 Qy 1427 GTTAAACCGTTTGGAGAGAGAGACGTTTTTGGGTAACTTTGATTAAGAGATGC 1486
 Db 5004 TTGAAACCATTGGAAAAAGAGTAGGTTTGGGTAACTTTGATTAAGAGATGC 5063
 Qy 1487 AAGAGCTTTTGAAGATGAAAAATGATCCATGTTACTTTTCAACAAATGCAATCT 1546
 Db 5064 TAGAGATTTCTAAGAGAGAGATGATATCATGCTTTTCCAAACAAACGTAATGT 5123
 Qy 1547 CATTAATTAGTAAAAATACCTTTAGCTTTCAAAATTTTACGTGATGAGAAACAGA 1606
 Db 5124 CAAAACTTGTAAAAAATACCTTTAGCTTTTAAAGTCTTTAGAAAAAAGAACAGACA 5183
 Qy 1607 TGTATATTTTCAATCTGTGCGGCGCTGTGCTGCCCTTTTCACTTGGAAACATAT 1666
 Db 5184 TGTATATATATCTGTGCGGCGCTGTGCTGCCCTTTTCACTTGGAAACATAT 5243
 Qy 1667 TGGAGCAAAAGACATTTATATGAGTATGATCGAGTTAATTAATCTACATTAATCTG 1726
 Db 5244 TGTGTGTAGACCGTTTATATAGAGTTTCGACAGATAGATTAACCACTTTGACAG 5303
 Qy 1727 AAAATAGTTTATCCGTAAACAGATATTTTATTTTGTTCAGTGGAGAAATGAAAGAT 1786
 Db 5304 AAAATAGTATCTGTATACAGATTAATTTATTTGTTCAGTGGAGAAATGAAAGAT 5363
 Qy 1787 ATATCTTAATCTATTAATCTTGGAGATATTTTATATGATTTTGTAAAGATGAACT 1846
 Db 5364 TATATCTTAATCTATTAATCTTGGAGATATTTTATATGATTTTGTAAAGATGAACT 5423
 Qy 1847 CATGAACAAGCTTTATGATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 1906
 Db 5424 CATGAACAAGCTTTATGATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 5483
 Qy 1907 ATTAACGACGAAATATTTATTTCAACAGATATTTGATTAATTTGATTAATTTGATTA 1966
 Db 5484 ATTAACGACGAAATATTTATTTCAACAGATATTTGATTAATTTGATTAATTTGATTA 5543
 Qy 1967 TATTAATAATTTTCTGATTAACAAAGAAATGAAACATATATTAACAAATGAAAGAT 2026
 Db 5544 TGTGTCAAAATTTCTCATATGATGATTAATGATTAATGATTAATGATTAATGATTA 5603
 Qy 2027 ATTTGCAAGGAGGCGCGCTACTTTTATGATTAATGATTAATGATTAATGATTAATGAT 2086
 Db 5604 ATTAACGACGAAATATTTATTTCAACAGATATTTGATTAATTTGATTAATTTGATTA 5663
 Qy 2087 TGTGTCTTAAGCAAAATGATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2146
 Db 5664 GTTGTCTTAAGCAAAATGATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 5723
 Qy 2147 AGAAGATTTTACAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 2206
 Db 5724 AAAAAAATTTGCAACCTGTATCCCTTGGCTTGGATGATTAATGATTAATGATTAATGAT 5783
 Qy 2207 AAAATATTTGAATTTTCTAAGCAAACTAATTTACATCAATTAATTTTGTGGA 2266
 Db 5784 GTGTGAAGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 5843
 Qy 2267 AGATTAATAAATAGTTGAAAAAT 2292
 Db 5844 AATTTGAAAAAATTTAGTGAAT 5869

RESULT 12

ADV81204_12/c

Continuation (13 of 23) of ADV81204 from base 120001 (Streptococcus agalactiae complete

WP Sequence split into 23 fragments LOCUS ADV81204 Accession ADV81204
 WP Fragment Name Begin End
 WP ADV81204_00 1 110000
 WP ADV81204_01 100001 210000
 WP ADV81204_02 200001 310000
 WP ADV81204_03 300001 410000
 WP ADV81204_04 400001 510000

WP ADV81204_05 500001 610000
 WP ADV81204_06 600001 710000
 WP ADV81204_07 700001 810000
 WP ADV81204_08 800001 910000
 WP ADV81204_09 900001 1010000
 WP ADV81204_10 1000001 1110000
 WP ADV81204_11 1100001 1210000
 WP ADV81204_12 1200001 1310000
 WP ADV81204_13 1300001 1410000
 WP ADV81204_14 1400001 1510000
 WP ADV81204_15 1500001 1610000
 WP ADV81204_16 1600001 1710000
 WP ADV81204_17 1700001 1810000
 WP ADV81204_18 1800001 1910000
 WP ADV81204_19 1900001 2010000
 WP ADV81204_20 2000001 2110000
 WP ADV81204_21 2100001 2210000
 WP ADV81204_22 2200001 2217924

Query Match 10.4%; Score 725.2; DB 13; Length 110000;
 Best Local Similarity 59.2%; Pred. No. 1.5e-88;
 Matches 1365; Conservative 0; Mismatches 903; Indels 38; Gaps 6

Qy 16 TGGCATTATTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75
 Db 89512 TAGCATGATTTCAACAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 89453
 Qy 76 ATGCTATTTTAATCG---TTCTGAAATTTTATCATATGATGATGATGATGATGATGAT 132
 Db 89452 CTCCCAATTTTAAACCAATTAAGATTTATGTTTGTCTTATGATGATGATGATGATGAT 89393
 Qy 133 TTTTATCTGCTATGCCAGTTGAATTTGATGATGATGATGATGATGATGATGATGATGAT 192
 Db 89392 TTTATCTTCTGATTTTATACAGACTTTTGTGAGTGTGCTATCTTGAAGATTTTAA 89333
 Qy 193 AAACATTTAATCTATGATTAATTTTGAATTTTCTTACGAGATCAATTTTGTG 252
 Db 89332 TGTGATTTGAATTAACGCTTTTACTATATTTTCAATCAATGATTAATTTTATTTTA 89273
 Qy 253 AGAATATTTTCCGATTTCAAGCGTGTGCGCTGATTTTCAATTAATTAATCTGTT 312
 Db 89272 AAAACCTTTTCAACAGACGATTTCTTTTATCTTTTATGATGATGATGATGATGAT 89213
 Qy 313 TGTATCTATTTAATGATTAATTTAAGATTTAAGATTTAAGATTTAAGATTTAAGATTT 372
 Db 89212 TATGATCTATGATTAATTTCAATTTTAAATATATGAAATATTTCTTACGTAAGTTT 89153
 Qy 373 TCTATCAAAAAAGAGATTTCTAATTTTCAACGCGCTGACGATGGAATATGCAATTT 432
 Db 89152 CACGATTAACCAAGTGTGTTGATTAACGAATTAAGATTTCTTATCAAAAATGACCTTT 89093
 Qy 433 TATTTGAATCAATTAACAAATTTCAAAAAATCTTTGCAATTTGATTTAGTACAG 492
 Db 89092 GGAATTAAT---ACGACCAATTAATATATGCTGTCTGATCTTCTGAAAG 89036
 Qy 493 AATATGATTAATTAATTTAT 552
 Db 89035 ATTTGATTAATTTGAAACATTAATCTGTTAAGATTAATTAACAAAGATCTTTAC 88976
 Qy 553 TTTCAACAGGAGAGTGTGACACGCTTTTATTAATCTTCAACAGTGAATTTTAGAGC 612
 Db 88975 AGTTAACGCTGCTTAATCTGTTGATCAAGCTTTTATTAATCAATCCATTTGAATTT 88916
 Qy 613 TAAAGCA-----TTGTTTCAATTTGATTTGATTTAGTATGATTAAGCGTTG 663
 Db 88915 AATTAACCAATTAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 88856
 Qy 664 ATATTAATTTATCTGCTTTTATCTGCTGTTGAAAAAGAAAAATCCAACTGCTAGAGAC 723
 Db 88855 ATGTAGAGCACTTACCTTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 88796
 Qy 724 ATAGCATTTAATTTTCCAAATTTTATTAAGCTTATGATTAATCAATGATGAAAGAC 783

```

Db      88795 ATAGTGTATATCATATCTATGAAATCTATATAAATATAGTACCTTATAGCAAAACGAT 88736
Qy      784 TTTTGGATTTACTCGAGCGGAGTAGTCGGGTAAATTAATTTGGTAAATGTTCTATTTTGT 843
Db      88735 TTTTGGATTTACCGGGTGTCTATTAATAGGTTTGTCTATATGTGCAATTTGGCAATTTTTC 88676
Qy      844 TAGTTCGAATTTATTCGTAGAGATGTGTGACCGGCTATTTTGTCTGAGAAACAGTTTGAC 903
Db      88675 TAGTTCGAATTTACGAAATAGATGTGTGACCGGCTATTTTGTCTGAGAAATAGTGTGTC 88616
Qy      904 AGAATGAGCGCATTTTATTAATCTTACATGATTCAGATTCATATGTTGATGCTGAGGAC 963
Db      88615 GTATGTGTAGATTTTATGATTTCTATATAATTCAGATCAATGCGAGTAGTAGCAACAA 88556
Qy      964 GCAAAAAAAGCTTGTCTACGCCAAACCAAGATGCAAGGGGTGGATTTTAAATGGAA 1023
Db      88555 TTAAAGAAATTTATTAATGTTTCAATCAATCAATGACAGG-----CTAATGTTAAGTTAA 88501
Qy      1024 AAACGATCTAGATTTACTCAATTTGACATTTTCATACGCAAAACAGTTTATAGCAGT 1083
Db      88500 AGATGATCTTAGAATTTCTAAATAGGAAATTTATTCG-AAAAAGATGATGAGT 88442
Qy      1084 TACCAGATTTTAAATGTTTAAATTTGCGATATGAGTCTAGTTGTTACAGTCCACTTA 1143
Db      88441 TGCCTCAATCTATATGTTTAAAGGTGATGTAGTTATGAGAAACCGCCCTCCA 88382
Qy      1144 CAGTTATGAATTTGAAAAATATATCCCTGTGCAAAAGAGCATTTGAGTTTAAACAG 1203
Db      88381 CAGTTATGAATTTGAAAAATATATCAACGAAAGGACGCCCTTATGTTTAAACAG 88322
Qy      1204 GATTTACAGTCTCTGCGAGGTAGTGTGCTGTATATATACAGACTTCGACGAGT 1263
Db      88321 GAATCTGTGTTGTGGCAATATCTGTGTAAGAAATATATCACTGATTTGAGAAATG 88262
Qy      1264 TTGCGTTGACCTTAGCATATCATTTGATATGATCTGTGTCAGATATTAATTTAT 1323
Db      88261 TAAAGTTAATGTTTCAATATATCAATGATGATGTTATTTGTGATATTAATATTC 88202
Qy      1324 TAAAGACAGTGAAGTTTATTTGTGAGAGAGGAAATGATAAGTA----- 1372
Db      88201 TCCTAACACTTAAGAGTGTCTTCTGTGACAGGTGCTAAGTAAAGGTTTGAAG 88142
Qy      1373 -----TATGAAATTTGTTGGTGTGCTTCTGAGGGGACATTTGACTCTGATTT 1426
Db      88141 GAATATATATAAATTTGTCTGTGTGTTCAAGTGTGTGATCTGACACATTTGAACCT 88082
Qy      1427 GTTAAACCGTTTGAAGAGAGAAAGACGTTTGGTAACTTTGATTAAGAAGATGC 1486
Db      88081 TTTGAACCCATTTGGGAAAAAGAAAGATAGGTTTGGTAACTTTGATTAAGAAGATGC 88022
Qy      1487 AAGAAGCTTTTGAAGATGAAAAATGATATCCATGTTACTTTCGCAAAATGCAATCT 1546
Db      88021 TAGAGATATCTAAGAGAGATGATATCATTTGCTTCTTCGCAAAACGTAATGT 87962
Qy      1547 CATTAATTTGAGAAAAATCTTTCTAGCTTTCAAAATTTTACGTATGAGAAACGAGA 1606
Db      87961 CAAAAACTTGTGTAATAAATATCTATCTAGCTTTTAAAGTCTTGAAGAAAGAACGAGA 87902
Qy      1607 TGTATATATTTCAATCTGTGTGCGCGGTGTGTCTTCTTTTAACTCGAAAACTATT 1666
Db      87901 TGTATATATATCATCTGTGTGCGCTGTAGAGATCAATCTTTTATATGTTAAGTATT 87842
Qy      1667 TGGAGCAAAAGCATTTATATTTGAAGTATTTGATCGAGTTATTAATCTACATTAATGG 1726
Db      87841 TGGTTGTAGAGCCGTTATATAGAGGTTTTCGAGAGATGATTAACCAATTTGACAGG 87782
Qy      1727 AAAACTAGTTATCCGTAACAGATTTTATTTTGTTCAGTGGGAGAAATGAGAGGT 1786
Db      87781 AAAATTTGTGTATCTGTATACAGATTAATTTATTTTGTTCAGTGGGAGAAATGAGAGGT 87722
Qy      1787 ATATCTTAATCTATTAATCTTGGGAGATTTTAAATGATTTTGTAACTGATGAAC 1846
Db      87721 TTATCTTAGGCAATTAATTTAGAGAGATTTTAAATGATTTTGTCACTGATGGGACA 87662

```

```

Qy      1847 CATGAAACAGTTTATCGATGATTAAGATTTGATTTTGAAGAAATGGAAGT 1906
Db      87661 CATGAAACAGTTTCAACCGTCTTATTAAGAAATGATGATTAAGAAAGGACAGTGTCT 87602
Qy      1907 ATTAACGCAAGAAATTTTATTAACAAACAGATATTTCTGATATATTCAGAAATATGCAAG 1966
Db      87601 ATTAATGCAAGAAAGTTTCAATTAACCGGTTTCTAGACTTTGAACCTCAGAAATGTGAG 87542
Qy      1967 TATAAATAATTTCTGATTTACAAAGAAATGCAATATATTAACAAATCAGAAATGAGT 2026
Db      87541 TGTGTAATAATTTCTCTCATATGATATGAATCTTTCATGAAAGAAAGCTGAGATGTT 87482
Qy      2027 ATTGCAACGAGGCGCGCTTCTTTATGAAATTCATATCCAAAGAAATTAATTA 2086
Db      87481 ATCAACATGCGGCGCTCAGCGACGTTTATGTGATTTTCTTTTGGAAATTAACAGTC 87422
Qy      2087 TTGTTTCTTAGCAAAAAAGATGATGAAACATGTAATGATCATCAATGATGATTTGTA 2146
Db      87421 TTTGTTCTTAGGAGAAAGCATTTTGTGAAACATATCATGATCATCAATATTTT 87362
Qy      2147 AGAAGATTTTACAGATTAATTAATTTTATTTATGAAATATATAGATTTGTTGAA 2206
Db      87361 AAAAAAATTTGCCACCTGATTCCTTGTGATTTGAAGATGATGACTTGCAGAA 87302
Qy      2207 AAAATTTATGAAGTTTCTAAGCAACTTAATTTTATTAATTAATTTTGTGAA 2266
Db      87301 GTTGTGAAAGAAATATAGCTACAGAAATATATCAGAAATATATATATGTTGTAT 87242
Qy      2267 AGATTTAAACAAATAGTTGAAAAAT 2292
Db      87241 AAATTAAGAAAAATTAATAGTGAAAT 87216

RESULT 13
ADO40235
ID ADO40235 standard; DNA; 25020 BP.
XX
AC ADO40235;
XX
DT 15-JUL-2004 (first entry)
XX
DE S. agalactiae capsular polysaccharide synthesis (cps) gene.
XX
KM Group B Streptococci; GBS; capsular polysaccharide synthesis; cps;
XX detection; ds; gene.
XX
OS Streptococcus agalactiae.
XX
FH key
FH misc_binding
FT 153..174
FT /tag= a
FT /bound_moiety= "S. agalactiae cps gene specific probe"
FT 546..581
FT /tag= b
FT /bound_moiety= "S. agalactiae cps gene specific probe"
FT 10100..10121
FT /tag= c
FT /bound_moiety= "S. agalactiae cps gene specific probe"
FT 10176..10198
FT /tag= d
FT /bound_moiety= "S. agalactiae cps gene specific probe"
FT 10219..10242
FT /tag= e
FT /bound_moiety= "S. agalactiae cps gene specific probe"
XX
XX US2004009574-A1.
XX PN
XX 15-JAN-2004.
XX PD
XX 09-JUL-2002; 2002US-00192280.
XX PF
XX 09-JUL-2002; 2002US-00192280.
XX PR
XX 09-JUL-2002; 2002US-00192280.

```


Query Match	Best Local Similarity	Score	DB	Length
Matches 1389; Conservative	58.9%	10.3%	DB 12;	25020;
	0;		Mismatches 926;	Indels 42; Gaps 7;
XX (DATP/)				
PA (SHAH/)				
XX (SHAH/)				
XX Data Gupta N, Shah K;				
PI WPI; 2004-090471/09.				
XX GENBANK; AF020896.				
DR Novel oligonucleotide probes for detecting group B Streptococci e.g.				
XX Streptococcus agalactiae in samples.				
PT Example 1; Fig 1; 28bp; English.				
XX The invention relates to compositions and methods for detecting group B				
CC Streptococci (GBS) Streptococcus agalactiae capsular polysaccharide				
CC synthetis (cps) gene in sample. The invention is useful for detecting S.				
CC agalactiae in a sample. The present sequence is S. agalactiae cps gene.				
CC This sequence is used to illustrate the method of the invention.				
XX Sequence 25020 BP; 8592 A; 3630 C; 4594 G; 8204 T; 0 U; 0 Other;				
16 TGGCATTTATGATATGATAGACAGTTGCAATTTCTGCAATCTTAAACAATATACCA 75				
5154 TAGCGATGATTTCAACAGTTGGTTTATTTTCTGCAAGTTGACATTAACATTAATTA 5213				
76 ATGCTGATTTAAATCGTTCTGGAATTTT---TATCATATATGATGTTTATTTTGA 131				
5214 CTCCCAATTTTAAAGCAATTAAGATTTATGTTGTTCTAATGATACATTAATATGTC 5273				
132 TTTTATATATCGATGATGACAGTTGAAATTTAGATATAGAGTAATCTGATAGATTGAA 191				
5274 TTTTATCTTTGATTTTATACAGACCTTTTGAGTGTGCTATCTTGAAGATTAA 5333				
192 AAAACATTTAACATATGATATATTTTGCAATTTTCTTACGGAGATATCAATTTTGTG 251				
5334 ATGATATGAATATACAGCTTTTACATATTTTCAATATCAATGATCATATTTTATTTT 5393				
252 GAGAAATATTTGCACTTCAAGACGTGTCGGTATTTTCAATTAATTAATCTGTT 311				
5394 AAAAATCATTTTAAACAACAAGCACTTTCTTTTCTTTTATGCTATTAATTCGAT 5453				
312 TTGGATATCATTTTAAAGTAAATTTAGAGAGTTTAAAGATAGAGTTTCTATTTTGCACA 371				
5454 TTAATGATATGATTAATCAATTTTAAATATTTATGCAAAATATTTCTTACGCTAATGTT 5513				
372 ATCTATCAAAAAGACATCTCTATATCAACGCTGAACGATGGAATAATATCAAGTT 431				
5514 TCACGAGATACCAAGTTGTTTGTATACGAATTAAGATTTCTTTATCAAAAATGACCTT 5573				
432 TTATTTGATCAATTAACAATTAATCTTGTGATGTTGATGTTTAAAGTAA 491				
5574 AGGATTAAT---ACGACATTAATTAATGCTGCTGACATCTTGAACCTCTGAAAAG 5630				
492 GAAATAGTAAATTAATTTATTCATTCACGCTCTATATTTCTGTGGAAGAGATATAG 551				
5631 GATTTGATTAATTTGAAACATTAACATCTGTTAAGATTAATTAACAAGATGCTTAACTTCA 5690				
552 TTTTCAACAAGAGATGTCGACACAGCTCTTATTAATCTTACGAAGGATTTTAAAG 611				
5691 GAGTTAACTGCTTAATCTGTTGATCAACGTTTATTAACATACCATGATTAATTTGCT 5750				
612 GTAAAGCAATTCG-----TTTCAATTTTGAATTTGTTAGTATTAATGATTAAGCTT 662				
5751 AAATACCAATTAACAAGATTAATTAATTAATGACATTTGAAGCAATGGGAGTATGTCATAGTT 5810				
663 GATATTAATTAATTCGTTTACGCTGTTGAAAAACAATAATCAATCTGCTAGCTGAC 722				
5811 AATGATGAGGCACTTAGCTTGTATATATATAGAGAAAGCAATTCCAACTTTTGAAGGA 5870				

QY	723	CATGCAATGTAACCTTTTCCACAAATTTTATATAGCCCTACTCATATCATATGAAACGA	782
Db	5871	TATGTGTTATTCAATATTCTTAGAATTTCTATTAATATATAGTCACTTATGCAAAACGA	5930
QY	783	CTTTGGATATATACCGAGCGGTAGTCGGGTAAATTATTTGGTATATGTTCTTATTTTG	842
Db	5931	TTTTTGGAATACAGGGGTCTATATATAGTTTGCTCATATGTGGCATTTGTGGCAATTTT	5990
QY	843	TTAGTTCATATATTCGTAGATAGTGTGACCGGCTATTTTGTCTCAGAAACGAGTTGGA	902
Db	5991	CTAGTTCACAAATCAGAAAAGATGTGTGACCGGCTATCTTTTCTCAAAAATGAGTAGT	6050
QY	903	CAGATGAGCCGATATTTCATCTTCACAGTTTGGATCGATGATATGTTGATGCTGAGGAG	962
Db	6051	CGTATGTGTAGAAATTTTTCATATCTTATTAATTCAGATCAATGACGATAGATGACGAACA	6110
QY	963	CGCAAAAAGACTTGCTCAGCCAAACACAGATGCAAGGGGTAGTATTTTAAATGGGA	1022
Db	6111	ATTAGAAAGATTTTATTTAGTTCAACATCAATGACAGG-----CTATGTTTAAAGTTAG	6165
QY	1023	AAAACGATCTTGAATTACTCCAAATTTGACATTTTATACGCAAAAACAGTTTGAACGAG	1082
Db	6166	AAGATGATCTTGAATTACTAAATATGGAATATTTATTCG-AAAAACAAGATATGATAG	6224
QY	1083	TTACCAAGTTTATATATGTTTAAATTTGCGCATATGAGTCTAGTTGGTACAGTCCACT	1142
Db	6225	TTGCCTCAATCTTAAATGTTTAAAGGTATGATGATTTAGTTAGGAACAGCCCTCC	6284
QY	1143	ACAGTTGATGAATTTGAAAAATATATCTCTGTGCAAAAAGACATGAGTTTAAACCA	1202
Db	6285	ACAGTTGATGAATTTGAAAAATATATTCATCGAAGACGACCCCTTAGTTTAAAGCA	6344
QY	1203	GGGATTAACAGTCTCTGGCAGTTAGTGTCGTATATATCAGACCTTGACGACGTA	1262
Db	6345	GGAATCACTGTTTGTGTGCAATATCTGTGAAATATATTACTGATTTTGAAGAATC	6404
QY	1263	GTTCCGTTGACCTTAGCATACATTGATTAATGACATCTGTGATGATTTAAATTTTA	1322
Db	6405	GTAAGTTAGATGTTCAATATATCAATGAATGCTATTTGTGATGATTAATTAAGTATAT	6464
QY	1323	TTTAAAGACAGGAAGTTGATTTGTGTAGAGAGGAAGTAGTAAAGTA-----	1372
Db	6465	CTCCTAACACTTAAAGTAGTCTTACCTGGGACAGGTGCTAATGAAGTAAAGTTTGAA	6524
QY	1373	-----TATGAAGTTGTTTGGTCGGCTTTGAGGGGACATTTGACTCACTGTATTT	1425
Db	6525	GGAATATATATAAAATTTGTCTGTGTGTTCAAGTGTGTGATCTAGACACATTTGAAC	6584
QY	1426	TGTTAAAAACGTTTGGAGGAGAGAACGTTTGGGTAACTTTGATTAAGAGATG	1485
Db	6585	TTTTGAACCATTTGGGAAAAATAAGTAGTTTGGGTAACTTTGATTAAGAGATG	6644
QY	1486	CAAGAGTCTTTGAAGATGAATAAATGATTCATGTTACTTTCCAAACAATGCAATC	1545
Db	6645	CTAGAGATATCTTAAAGAAAGATGTGATATCATCTTGTCTTTCCAAACAACCGTATAG	6704
QY	1546	TCATTAATTTTGTGAATAAATCTTTCTTAGCTTCAAAATTTTACGTAGAGAACGAC	1605
Db	6705	TCAAAAACCTTGTAAAAAATATCTATCTAGCTTTTAAGGTCCTTAGAAAAAGAAAGACAG	6764
QY	1606	ATGTTATTAATTTCAATCTGTGTCGGCCGCTGTGCTCCCTCTTTTAACCTGGAACATAT	1665
Db	6765	ATGTTATATATATCTGTGTGTCGGCTGTGAGACATCTCTTTTATATTTGTGAAGTAT	6824
QY	1666	TTGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTAATTAATCTACATTAATG	1725
Db	6825	TTGGTTGTAAGACCGTTTATATTAAGGTTTTCGACAGATATGATTAACCAACTTTGACAG	6884
QY	1726	GAAAACATGTTTATCCGCTACACATATTTTATTTGTTACGTGGGAGAAATGAAGAAG	1785
Db	6885	GAAAATTAATGTATCTGTAAACAATTAATTTATTTGTCAGTGTGGAGAAATGAAGAAG	6944

```

Oy 1786 TATATCTTAATCTATTACTTGGGAGTATTATTTTAAATGATTTTGTGTAACAGTAGAAGC 1845
    |||||
Db 6945 TTTATCTTAAGGCAATTAATTAGAGGAAATTTTAAATGATTTTGTGCAGTGGGAGC 7004
Oy 1846 TCATGAACAACAGTTAATCGATTGATTAATAAGATTTGATTTTATGAAAAAATGAGAA 1905
    |||||
Db 7005 ACATGAACAGCAGTTCAACCGTCTTAATTAAGAAAGTTGATTAATAAGGAGCAGGTGC 7064
Oy 1906 TATTAACCGGAAATATTTATCAAAACAGATATTCTGACTATTTCCAGAAATTTGCAA 1965
    |||||
Db 7065 TATTGATCAAGAGTGTTCATTCAACGGGTTACTAGACTTTGAACCTCAGAAATTTGTA 7124
Oy 1966 GTATATAAAATTTCTCAGTTTACAAAGAAATGCAATATATTTAACAAATCAGAAAGTAGT 2025
    |||||
Db 7125 GTGGTCAAAATTTCTCTCATATGATGATGAACTCTTCAATGAAAGAGCTGAGATTGT 7184
Oy 2026 TATTTGCCAGAGCGCCGCTACTTTTATGAAATTCATTAATCCAAAGAAAAACAAAT 2085
    |||||
Db 7185 TATCAACAATGGGGGCTCAGCGACGTTTATGAAATGACAGTTTCAAAAGGAAAAAATCTAT 7244
Oy 2086 ATTGTTTCTTACACAAAAAAAGTATGCTGAACATGTAATCATCAAGTAGAGTTGT 2145
    |||||
Db 7245 TGTGTTCTTACAGCAAGAACAGTTTGAGAGCATGTAATCAATCAGAGTGAGATTTTT 7304
Oy 2146 AAGAAGAAATTTTACAGATTAATATTTTATTTATGAAAAATATGATGATTGTTGA 2205
    |||||
Db 7305 GAAAGAGTATTTCTTGAATTAATGATTAATTTGAAATATCAGTAATTAAGAA 7364
Oy 2206 AAAAATTAATGAAGTTTCTTAAGCAAC--TAACTTACATCAATTAATTTTGTG 2262
    |||||
Db 7365 TATTATTAAGAAAAAAATATATCTATGTAAGTAATATCAAAACAAATGATTTTGTG 7424
Oy 2263 TGAAGATTAACAAATGTTGAAAAATTTATGAGATCAAGAAATGATTAATAAA 2322
    |||||
Db 7425 TTTCTCTTCAAAAAAGAACATTTCAATAAATAATGATTAATATTTTGTGAGAA 7484
Oy 2323 AAGATGATATTTGATA 2339
    |||||
Db 7485 AAAAATTTGAATTAACA 7501
    |||||

RESULT 14
ADM79774
ID ADM79774 standard; DNA; 2226 BP.
AC ADM79774;
DT 03-JUN-2004 (first entry)
DE Group B Streptococcus cpsB-cdsF-cpsF-cpsG partial consensus DNA sequence.
KW Group B streptococcus; GBS bacterium; cpsB; cpsF; cpsG; cpsI gene;
KM M gene; GBS infection; ds.
XX
OS Streptococcus sp. 'group B'.
EN WO2003025216-A1.
XX
PD 27-MAR-2003.
XX
PF 18-SEP-2002; 2002WO-AU001281.
XX
PR 19-SEP-2001; 2001AU-00007749.
XX
PA (MSYD-) WESTERN SYDNEY AREA HEALTH SERVICE.
XX
PI Fanrong K, Gilbert G;
XX
DR WPI; 2003-381495/36.
XX
PT Typing a group B streptococcus (GBS) bacterium for diagnosing the GBS
    infections in pregnant women, elderly or immunocompromised patients by
    analyzing the sequence of the regions in the cpsD, cpsE, cpsF, cpsG or

```

```

PT cpsI/M gene of the bacterium.
XX
PS Claim 2; Fig 1; 106pp; English.
XX
CC This invention relates to a novel method of typing a group B
CC streptococcus (GBS) bacterium which comprises analysing the nucleotide
CC sequence of one or more regions within the cpsD, cpsE, cpsF, cpsG and/or
CC cpsI/M genes of the bacterium, where the regions comprise one or more
CC nucleotides having sequences that vary between types. The method is
CC useful for preparing a composition for serotyping and/or subtyping a GBS
CC bacterium for diagnosing GBS infections in pregnant women, elderly and/or
CC immunocompromised patients. The present sequence is that of a consensus
CC DNA sequence for the 3' end of the cpsD-cpsE-cpsF and the 5' end of the
CC cpsG sequences of group B Streptococci which is related to the method of
XX the invention.

SQ Sequence 2226 BP; 767 A; 293 C; 417 G; 749 T; 0 U; 0 Other;
Query Match 9.3%; Score 649.2; DB 11; Length 2226;
Best Local Similarity 59.5%; Pred. No. 2.8e-78;
Matches 1223; Conservative 0; Mismatches 793; Indels 38; Gaps 6;

Oy 16 TGGCATTTATTTGATATGATAGAGTTCGCAATTTCTGCAATCTTAACAAGTCATATACAA 75
    |||||
Db 182 TAGCATGATTCAAACAGTTGTGTTTATTTTCTGCAAGTTGACATTAACATTATTA 241
Oy 76 ATGCTGATTAATGCG--TTCTGGAATTTTATGATATGATGATGTTTCAATTTTGCAT 132
    |||||
Db 242 CTCCTCAATTTTAAAGCAATTAAGATTAATGTTTCTATTTGATATGATTAATATTTGCT 301
Oy 133 TTTTATATCTGATGCGAGTGAATTTGAGTATGAGTAAATGATAGAGTTGAAA 192
    |||||
Db 302 TTTATCTTCTGATTTTAAACAGACTTTGAGTGTGCTATCTGAAAGTTTAA 361
Oy 193 AAACTTTACTATATGATATATATATTTGCAATTTTCTTACGCGAGATCATTTTGTG 252
    |||||
Db 362 TGGATTTGAATTAACAGCTTTTCTATATTTTCAATCAAGTTCAATTTTATTTTAA 421
Oy 253 AGAATTAATTTGCACTTTCAAGACGTTGCGGTGATTTTCAATTAATAAATCTGCTT 312
    |||||
Db 422 AAAACCTTTTAAACAGACAGACTTCTTTTAACTTTTATGATGAATTCGATTT 481
Oy 313 TGGTATACCTATTTAAAGTAATTTAAGACAGTTTAAAGATGCTTCTATTTTGCACA 372
    |||||
Db 482 TATTATATCTATGATTAATTTTAAATATATTCGAAATATTTCTTACGCTAAGTTT 541
Oy 373 TCTATCAAAAAAGACATTTCTAATTAACAAGCGTGAACAGATGGAATAATGCAAGTT 432
    |||||
Db 542 CACGAGATACCAAGTTGTTTGAATTAAGAAATTAAGATTTCTTATCAAAAATGACCTTTA 601
Oy 433 TATTTGAATCATAAACAATTCAAAAAATCTTGTGCAATTTGATGTTTAAAGTACAG 492
    |||||
Db 602 GGAATTAAT--ACGACCAATTAATATATGCGTGTCTGATCTTGACCTCTGAAAAG 658
Oy 493 AAATGATTAATTAATTTATATCAATCCGCTATTTATTTCTGGAAGAAGCTATAGAGT 552
    |||||
Db 659 ATGTTATGATTTGAAACAATACTCGTTAAGATTAATAACAAGATGCTTTACCTTCAG 718
Oy 553 TTTCAACAAGGAAAGTGTCGACCAAGCTTTAATAATCAATCAAGTGAAGTTTAA-- 609
    |||||
Db 719 AGTTAACTGCTTAACGTGTGACAGCTTTTATTAATCAATCCATTTGAATTTTGGTA 778
Oy 610 -----ACGTAAGCAATTCGTTTCAATTTGAGTTGTTAGTATGATGATGAGGTTG 663
    |||||
Db 779 AATACCAATTAACAAGATTTATTAATGAACATTAAGCAATGGAGTGATGTCATAGTTA 838
Oy 664 AATTAATTCATTCGCTTTTACTGCGTTGAAAAACAATAAATCAACTGCTAGGTGACC 723
    |||||
Db 839 ATGAGAGGCACTTACCTTTGATTAATATGAGAAAAAGCAATCAAACTTTTGAAGAT 898
Oy 724 ATAGCATTTGATTTTTCACAATAATTTTATAGGCTGATCATCATGATGAAGACAG 783
    |||||
Db 899 ATAGTGTATTAATATTTATGAAATTTCTAATAATTAATGATGACCTTATAGCAAAACAT 958
    |||||

```


PA (USDA) US DEPT OF AGRICULTURE.
 XX Trempey JB, Knoshaug EP, Sandine WE, Ahlgren JA, Dietken KP;
 XX WPI: 2001-488889/53.
 DR P-PBDB; AAB47416, AAB47417, AAB47418, AAB47419, AAB47420, AAB47421,
 DR AAB47422, AAB47423, AAB47424.
 XX
 PT New bacterium useful in pharmaceutical formulations, food products and
 PT beauty cosmetics, comprises characteristics of *Lactococcus lactis*
 PT cremoris Ropy 352.
 PS
 PS Disclosure; Page 50-58; 73pp; English.
 CC This sequence represents a fragment of the EPS plasmid derived from *L.*
 CC *lactis* subspecies *cremoris* Ropy352. The EPS plasmid is about 32 kb in
 CC size and encodes at least 13 active genes. The enzymes encoded by these
 CC genes allow the bacteria to produce an exopolysaccharide, designated
 CC EPS352. When EPS352 is expressed in or added to milk, it imparts highly
 CC desirable sensory characteristics to the milk, including making the milk
 CC very thick, with a very smooth mouth-feel, and slightly sweet with an
 CC obvious chewable-bite. Open reading frames (ORF's) M and N show homology
 CC to glycosyltransferase involved in EPS352 biosynthesis. *L. lactis*
 CC cremoris Ropy352 is deposited with the USDA-ARS-NCAUR-NRRL as deposit
 CC accession number NRRL B-30229. EPS352 is useful for thickening a liquid
 CC selected from milk, a milk-based liquid, a whey-based liquid, a soy-based
 CC liquid, and a fruit-juice. It is also useful as an additive in
 CC pharmaceutical products, beauty care products and coating agents. It is
 CC indicated, schematically, in the specification that there is an
 CC additional ORF between those encoding EpsB and EpsO. This encodes EpsF
 CC (AAB47425). However, this ORF is not indicated on the actual DNA sequence
 XX
 XX
 SQ Sequence 6850 BP; 2487 A; 965 C; 1206 G; 2192 T; 0 U; 0 Other;
 Query Match 6.1%; Score 428.4; DB 4; Length 6850;
 Best Local Similarity 58.9%; Pred. No. 1.le-48;
 Matches 849; Conservative 0; Mismatches 581; Indels 12; Gaps 6;

861 AGAGATGCGGACGCGCTATTTTGTCTCAGAAAAGATTGACAGATGAGCGCATATTT 920
 3518 AAGAGATCAAGGGCCAAATGTTCTATTAACAAAAGCGTATGTAAATGTAAATTTT 3577
 921 ACATTTACAAAGTTGCGATGATGTAATGTTATGCTGAGAGCGCAAAAAGATTGCTC 980
 3578 TATATTTTGAATTTTGAACAATGATCTTAATGCCAGCA--GTATCTGAAGCTTAATC 3635
 981 AGCCAAAACCAATGCAAGGCTGGTATGTTTAAATGGAAAAAGATCTTAAGATTA 1040
 3636 CAGATGTTAAAGCTGCTTACCATGCCAAACGCAATAGCTAAGAAAGATCCAGGGTAA 3695
 1041 CTCCAATTTGCAATTTGATACGCAAAAAGATTGACAGATTACAGAGTTTATAT 1100
 3696 CGAAGATTGGCTCATTTATTAAGACGACATCAATT-GATGAATCTGCCAATTTTATCAT 3754
 1101 GTTTTAAATGGCAATAGATGATGTAATGTTGTAACGTCACCTACAGTTGAATTTGAA 1160
 3755 GTTCTTAAAGGGATATGTCATTAGTGTGTCGAAGCAATTCGCTTTTGAAGCAAA 3814
 1161 AAATATATCTCTGGTCAAAAAGACGATGAGTTTAAACCGAGGATTAACAGCTCTGG 1220
 3815 GAATATGGAAGCGCTCGCTACTTACTCATGTCAAACGAGAAATCACCTGGTTATGG 3874
 1221 CAGGTTAGTGTGTAATATATACAGACTTCAGACAGATGTTGCGTTGACATTAGCA 1280
 3875 ACGACATGTCGTCAAGTAAAGTCTTTTTCCTCAAGAGAGAT---TTAGAACTCTAT 3931
 1281 TACATGATATATTGAGCTATCTGCTAGATTTAAATTTTAAAGACAGTAAAGTT 1340
 3932 TATCTCAATGACATAGACCAAAAATGATATCAAGCTTCTAGATCTCAAGATTTGACAA 3991
 1341 GTATTTGTTGAGAGAGGAAGTAAGTAATGAAGTTTGTGGTGGTCTTCA 1400
 3992 AGATTTAAGGATGCGACGATATTAAAA---ATGAAAAATAGCATTAGTAGTTCCAGC 4048

QY 1401 GGGGACATTTGACTCTACTTGTATTTGTTAAACCGTTTGGAAAGAGAACGTTT 1460
 DB 4049 GGTGGCCATTTGACACACCTGTATTTGTTAAAAAGTTTGGAAAAAGAGATGATTT 4108
 QY 1461 TGGGTACATTTGATTAAGAGAGATGACAGAGTCTTTTGAAGATGAAAAATGATCCA 1520
 DB 4109 TGGGTACATTTGATTAAGAGAGATGACAGAGTCTTTTGAAGATGAAAAATGATCCA 4168
 QY 1521 TGTATCTTTCACAAATTCGCAATCTCATTAATTTAGTGAATAATCTTTAGCTTTC 1580
 DB 4169 TGTATTTATCCACAAATAGAAATGTAAAAAACAAGATTAATAATCCATCTTGCATTT 4228
 QY 1581 AAAATTTACGTATGAGAAAACCAAGTCTTATTTATCATCTGGGCGCGCTGCTGTC 1640
 DB 4229 AAAATTTACGTATGAGAAAACCAAGTCTTATTTATTTAGAGTGTGCTGCGTACGTT 4288
 QY 1641 CCCTCTTTTACATGCAAAAATCTATTTGAGCAAGACGATTTATTTGAATATTTGAT 1700
 DB 4289 CTTTCTTTTGGTTAGTAACTATTCGTGCAAGACAGTCTATTTGAATAATTTGAC 4348
 QY 1701 CGAGTTAAATTAATCTTACATTAATCTGAAAACCTAGTTTATCCGTACAGATATTTTAT 1760
 DB 4349 CGATTCGATTAACCACTTAACAGAAAATTAAGTTTATCCAGTTACGATTAAGTTTATA 4408
 QY 1761 GTTCGTGGAAGAAATGAAGATATATCTTAATCTAATTAATCTGGGGATTTT 1820
 DB 4409 GTTCAATGGGAAGATTAATAAAGTTTAACTTAAGCAATTAATTTAGAGAAATTTTC 4468
 QY 1821 TAAATATTTTGTAAACAGTAGAACCTCATGAACAACAGTTTAACTGATTAATAAGAGA 1880
 DB 4469 TAAATATTTTGTAAACAGTTGAACTCATGAACAACATTTAATGACTCAATCAAAA 4528
 QY 1881 TTGATTTATGAAAAAATGAAAGTAAACCGAATAATTTATTTCAACAGATAT 1940
 DB 4529 TTGATGAACTTTGACGCAATGTTGAATGAAGACAGATGATTTCAATGCAATTTGGGTA 4588
 QY 1941 CTGACTATATTCAGATATTTGCAAGTATTAATAAATTTCTCAGTTTCAAAAGAAATGGAAC 2000
 DB 4589 CAACCTATGAACTTAATATTAATAATGGAATAATTTATTTGATATGACTATGAGAAA 4648
 QY 2001 AATATTTTAACAATGCAAGATTAATTTGCAAGGAGCCCGCTACTTTATGATAT 2060
 DB 4649 GATGATGATGAAAGGAGTACGATATTTATCTCATGCGGACATCTATATGCAAG 4708
 QY 2061 CATTTATCAAGGAAAAAACAATTAATTTTCTTACACAAAAAAGTATGTAACATG 2120
 DB 4709 TATTTCAACTAGTAAATTTCCGATAGTTTCCAGCGCAAAATGAAATTTGATGACATA 4768
 QY 2121 TAAATGATCATCA--GTAGATTTGTAAGAAATTTTCAAGATTAATTAATTTATTT 2179
 DB 4769 TAAATGATCATCACTTTGGTAAATGAAACAGGTTGTGAAAAAGGATATCTCATGATTT 4828
 QY 2180 --ATGAAATATATGATGATTTGTTGAAAAATTTATGAAAGTTTCTAAGCAACTACT 2237
 DB 4829 TGTGCAAGATGTTGGAACATTTCTGAAAAATTTATTTGTTCAAAATTTTCAGATACCT 4888
 QY 2238 TTACATCAATATTAATTTTGTGAAAGATTAATAAACAATAGTTGAAAAATTTATG 2297
 DB 4889 TACAAAAAATGTAATTCACACACATGAAATTCATTAATTTATTCAGTGTGAAATTTAC 4948
 QY 2298 AG 2299
 DB 4949 AG 4950

Search completed: December 25, 2005, 05:20:30
 Job time : 2599 secs

[illegible]

OY		2980	AAATTTAAAACAATTCTTACTGAAAAAATAAAAGTGTAATTTGTGGACGTAATGCATG	3038
Dn		747	AAATTATVAAATTTAT	688
OY		3040	TTCAAATTAATTTAA-----TAGCACCCGAATATTTATTTTTTAAGTACTTCGTTGAT	3094
Dn		687	TTTTTTTATTTAT	628
OY		3095	TATTTTATTTCCGAGCAAAAGAATGATGATTTTATTAATTTTATATGAATTTATTTAT	3154
Dn		627	AATTATVAAATTTATTTAAATTTATATATAAAATTAATNTNTTTATATAATTTAAAT	568
OY		3155	TCAT--ATAAAATTTTGAAGAACCTAGCTAATTTAAAAATGAAATTTTATTTGTTTTTA	3212
Dn		567	TNATTAATTAATTAATATATATATATATANTATATATATATATATATATATATATATTA	508
OY		3213	TTATGCTCATAATTANGTTTTGTTTCAGTAGTCAAATATGTTTGTGCAATTAATTTT	3272
Dn		507	TTATTAATAATTTTATTT	448
OY		3273	GAAAGATTATTTGGAGATTTTACGCCCATATATTTGGATTTTGGCAATAATGATTAAT	3332
Dn		447	TTTATTAATTAATTAATA	388
OY		3333	AAATTGTATTCATTTATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTTTACT	3392
Dn		387	TTATTAATTTATTTAAATATANNITATATATATTTTATATTAATTAATTAATTAATTA	328
OY		3393	TTTTTACTTTTATAGTATATCTGCATCTGATATATATTTTCAAAATGGAAAGATTTGTA	3452
Dn		327	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTATTA	268
OY		3453	TTTTTAGCACGACCCTTAT 3472	
Dn		267	TAATTAATAATTAATTAATATAT 248	
RESULT 2				
CG753083			1896 bp DNA linear GSS 24-OCT-2003	
LOCUS			Pd48-I-COI_za Ppa EcorI BAC library Pristionchus pacificus genomic,	
DEFINITION			genomic survey sequence.	
ACCESSION			CG753083	
VERSION			CG753083.1 GI:379771199	
KEYWORDS			GSS.	
SOURCE			Pristionchus pacificus	
ORGANISM			Eudaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;	
REFERENCE			Nediplogasteridae; Pristionchus.	
AUTHORS			1 (bases 1 to 1896)	
TITLE			Srinivasan,J., Sinz,W., Jeesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J. An integrated physical and genetic map of the nematode Pristionchus pacificus	
JOURNAL			Mol. Genet. Genomics 269 (5), 715-722 (2003)	
PUBMED			12884007	
COMMENT			Contact: Sommer RÜ Evolutionary Biology Max-Planck-Institute for Developmental Biology Spemannstr. 37-39, Tuebingen D-72076, Germany Tel.: 00497071601371 Fax: 00497071601498 Email: ralf.sommer@uebingen.mpg.de Class: BAC ends.	
FEATURES			Location/Qualifiers	
source			1..1896	

```
/note="The library was generated by a partial digest of
```

the genomic DNA with *Ec*RI and cloning into the BAC vector."

Query Match	2.4%	Score 167.4;	DB 10;	Length 1896;
Best Local Similarity	46.8%;	Pred. No. 1.3e-15;		
Matches 761; Conservative	0;	Mismatches 830;	Indels 35;	Gaps 9;

QY	2050	TTTTATGCAATTCATTAATCCAAAGGAAAAAAACAATTATGTGTTCTCAGACAAAAAAGATA	2109
Db	219	TNATATTTATTTNATATTTATTAANNAATTAATATATTTATATTTATATATATATATATATAT	278
QY	2110	TGCTGAACATGTAAATGATCATCAAGTAGAGTTGTGAAGAAATTTTACAGATATATA	2163
Db	279	TANATATATANNANATATTTANNANNAATTTTATTTATATTTATATATATATATATATATA	338
QY	2170	TATTTTATTTATAGAAAAATATAGATGATTGTTGAAAAAAT-TATGAGTTTCTAAGC	2228
Db	339	TATATATTTTATTAANNAATTAATATANNATATATTAATTAATATATATATATATATATTA	398
QY	2229	AAACGAACCTTAACATCAATATATATTTTTTTGGAAAGATTAACAAATAGCTGAAG	2288
Db	399	TTATATAATNATTTTATATATATTAATTAATATATATATATATATATATATATATATATA	458
QY	2289	AATTTAATAGAGATCAAGAAAAATGAAATATATAAANAAGATGCAATTTGATATAGCTTAT	2348
Db	459	AAATATATTTAAT	518
QY	2349	CATAATTTTCTCAGATTTTACTGAGAGGAGATCAGATATATATCATCTTCTCAGAG	2408
Db	519	AATTTTATATTTTATTA	578
QY	2409	AATGACACCATAGTTCCTCAGAAATCCGTAATATATTTTAAATATCTCAGAGTT	2468
Db	579	ATTATATATAAANNAATTTATTTATATATATATTTTAAATTAATATATTTATATATATAT	638
QY	2469	TATATGTTGAATTTTACAAAGATGAGCAAAATATTAAGAAATAGATATATGACAGC	2528
Db	639	ATATTTNNAAT	698
QY	2529	TTAAATGTTACAGATTATTTCTTAATATATCAGAAAAACATATGA--TAAATGATCTGT	2588
Db	699	ATATTTTAAATTTTATATTTTAAATTTTATATTTTAAATTTATTAATTTATATATATATA	758
QY	2587	TAGAATTTTATTAAGATGTATCCAGCTTTTGAATATCATATTTTACAAAGATGTGTTAT	2648
Db	759	TAAATTTAATATATTTTAAATATATTAATAAATATTTATATTTTAAATTTTATATATTTT	818
QY	2647	TGATAGATATAAAAAACATGCTCTAAGATAAGATTGGTCTAAATTTGGTTCGCTCCA	2706
Db	819	TNAT	878
QY	2707	CATGATTTTGTGGCAATCTTTTATCAAAATGA--AAGAAAACAGCTTATTTATTTAAGTA	2765
Db	879	ATTATATTTTATATATATATATATATATTTTAAATATTAATTTTATTTTAAATTAATA	938
QY	2766	ATCTAATATGTCAGATG-----AAGTATTTATACAGCAATTTATAGAAAAATGAATTT	2820
Db	939	AATTTATTTTAAATATATTTTAAATTAATATATATATATATATATATATATATATATAT	998
QY	2821	TCAATATGATTTATCTAATATAGAAATTTAAGATATATAAAGTGTG-----	2865
Db	999	AAAATTTATATTAATTTTATTTATATATATAATATATATTAANNAATATATATATATAT	1058
QY	2866	-----AAAAATGCAACATCTTCTCCTATGTCTTTACAGATGATCTATGATGAATGGTA	2922
Db	1059	ATTAATATTAATTAAT	1118
QY	2923	AATGCAAGAAATTTAGCTTT---TTATTTGCTAGAAAGTAAAAATGAAAAATTAATCT	2979
Db	1119	TATATATATATATTTTATTTTATATATTTATTTTATTTATATATATATATATATATAT	1178
QY	2980	AAATTTTAAAGAAATTTACTTAAAAAATTAATAATAGTGAATTT---TGAGAGTATATGTA	3036

[illegible]

```

Email: ralf.bommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1. .1592
FEATURES
source
```

ORIGIN

[illegible]

OY	2892	CTTTACGATGATTCATTTGATGAAATTCGAAAATGCAAGAAATTTAGTCTTTTATATTCG	2951
Db	692	TATTTAAAAAT	633
OY	2952	TAGAAAGTTAAAAATAGAAAATTAATCTAAATTTAAAGAAATTAATTAACAAAAATVAAA	3011
Db	632	AATTAATTAATTTTANNTATATATAATTAATAATAAAAATAAAAATAAAAATAAAAA	573
OY	3012	TAGTTGATTTTGAGAGATGATGATGATTTTAAATATTTAAATATTTAAATATGACCGGAATATTT	3071
Db	572	ATTTNNNTTNNNTTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT	513
OY	3072	ATTTTAAAGTCTCGGTGATTAATTTTAAATTCAGAGCAAAAGATATTTTATTA	3131
Db	512	TTTTNNNTTTTTTTTTTTTTTTTTTAAATNNNTT-TNNNTNTNTNTTTTTTTTTTTTTTT	454
OY	3132	ATTTTATGATTTAAATTTTATTTTCATATATAAATTTTGGAAAATGAAGTAAATATTA	3191
Db	453	TTTNTTTTTTTTTTTTTTTTTTNTTNNATATAAANNTTTTTTTTTTTTTTTTTTTTTTTT	394
OY	3192	AATGAAATTTTATGTTTTTATATATGCTATATATATGTTTGTTCCAGATGTCACAGT	3251
Db	393	TT	334
OY	3252	ATGTTTGTAATAAATTTTGAAGAATTAATTCAGATTTTACGTCCCATATTTTG	3311
Db	333	TT	274
OY	3312	ATTATTCGAATTAATGATTAATATATTTGATTCATTTATAATATGATTAATAAAAAATTA	3371
Db	273	TT	214
OY	3372	AAAAATAGTATCTTTTATGTTTTTATGTTTTTATAGTATATCTGCATCTGATATATAT	3431
Db	213	NNNTTNN	154
OY	3432	CAAAATGGGAAGAATATGATTTTTTG	3459
Db	153	NNNNNNNNNNNNNNNNNNNNNTTTNNNG	126

```

RESULT 4
LOCUS      CG749499
DEFINITION CG749499                                1348 bp      DNA          linear      GSS 24-OCT-2003
            P043-4-A06.za Ppa EcORI BAC library Pristionchus pacificus genomic.c.
            genomic survey sequence.
ACCESSION  CG749499
VERSION     CG749499.1  GI:37970425
KEYWORDS   GSS.
SOURCE      Pristionchus pacificus
            Neodiplogasteridae; Pristionchus.
ORGANISM   Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE  1 (bases 1 to 1348)
            Sriivisan,J., Szinz,W., Jeesse,T., Wiggers-Perebolle,L., Jansen,K.,
            Bunjer,J., van der Meulen,M. and Sommer,R.J.
            An integrated physical and genetic map of the nematode Pristionchus
            pacificus
            Mol. Genet. Genomics 269 (5), 715-722 (2003)
JOURNAL    12884007
PUBMED     Contact: Sommer R.J
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497072601371
            Fax: 00497072601498
            Email: ralf.sommer@cuebingen.mpg.de
COMMENT    Class: BAC ends.
FEATURES   Location/Qualifiers
            source          1..1348
                        /organism="Pristionchus pacificus"
                        /mol_type="genomic DNA"

```

/strain="California"
/db xref="taxon:54126"
/clone lib="Ppa Ecori BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN

Query Match 2.3%; Score 158; DB 10; Length 1348;

Best Local Similarity 47.1%; Pred. No. 3.7e-14; Matches 575; Conservative 0; Mismatches 631; Indels 15; Gaps 3;

2928 AAGAAATTTAGTTTATTTGCTAGAAAGTAAAAATAGAAATTAATCTAAATTTAA 2987
128 AAAAAATTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTAA 187
2988 AGAAATTTACTTAAATAAATAAGTGTATTTTGAGAGTAATGATGTTAAATTA 3047
188 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 247
3048 TTTAATATGACCGGAGATATTTTATTTTAAAGACTGCTGATTTATTTATCCGA 3107
248 TTTTAAATATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 307
3108 GAGCAA--AAGTATGATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTA 3164
308 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 367
3165 TTTTGGAAAATAAGCTAATATTTAAATAAGAAATTTATTTGTTTATTTATGCTTAA 3224
368 TTTTATTTTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTT 427
3225 TTTATTTTGTTCAGAGTACAGAGATGTTTGTGAATTAATTTTGAAAGATTATTT 3284
428 TAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 487
3285 GCAGATTTTACCTGCCATTAATTTGATTAATTCGATTAATGATTTATTTGATTTCA 3344
488 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 547
3345 TTTTATTAATTTGATTTTATTTAAATAATTTAAATAATTTTATTTTATTTTATTTT 3404
548 TATTTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 607
3405 TTTAGATATCTGATTTATTTATTTCAAAATGGAAGATATGATTTTATTTAGACAGA 3464
608 TTTTATTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 667
3465 CACCTTATAGACTAGACTATCTTATTAACAGGCGTCAAAACAAGTTGGCTTATG 3524
668 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 727
3525 AACATCTCTACGTTAATATACACATTAATATGTTCAATTCGTTAATCTTTGCACTT 3584
728 NTTTTAAATTTTAAATAATTTATTTATTTATTTATTTATTTTATTTTATTTTAT 787
3585 ATAAAAAATAAATGCAACAATTTTCTGTGCTGCTCTTTATACGATCTATTTA 3644
788 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 847
3645 AGTGATCGAGATTTGATTTATGCTGCAATATTTATTTATTTATTTATTTATTT 3704
848 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 907
3705 TATATAGTGAATAATTTGCTGATTTAAATAAGCTAATATTTTATTTATTTATTT 3764
908 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 967
3765 ATTTATTTAAATCTGAATGCTTTACCAAGAAATTT-----TTGGCTGTTATA 3813
968 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1027
3814 ATCTAGAGATCAAGTACGAAGCTAGATTTTATTTATTTATTTATTTATTTATTT 3873

1028 TTTTATTAATAATATTTAAATTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1087
3874 TATTTAGAAAACATATTTTATTTTGGATATGGAATATCCGAATATTCAGTACGGA 3933
1088 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1147
3934 GGCTCGAGATCATTCAGCTATATATCATTTTATTTTATTTTATTTTATTTTATTTT 3993
1148 AAAAAATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 1207
3994 TTTTACTGATGTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 4053
1208 TTTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 1267
4054 CAGCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 4113
1268 TATATATTAATAT-TATTTAAATTTATTTTATTTTATTTTATTTTATTTTATTTAT 1326
4114 TTTATTTATTTATTTAGTACTAT 4134
1327 TATATATTTTATTTATTTATTTAT 1347

RESULT 5

CF238805 1626 bp mRNA linear EST 05-AUG-2003

LOCUS AGENCOURT_15099447 NICHD_XGC_bmb6 Xenopus tropicalis cDNA clone

DEFINITION IMAGE:6995950 5', mRNA sequence.

ACCESSION CF238805

VERSION CF238805.1 GI:3442013

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 1626)

NIH-MGC

http://imgc.ncl.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: c9apds-remail.nih.gov

Tissue Procurement: Robert M. Grainger

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://imgc.lnl.gov

Plate: L1M41680 row: 9 column: 21

High quality sequence start: 71

High quality sequence stop: 316.

Location/Qualifiers

1. 1626

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/db_xref="taxon:8364"

/clone="IMAGE:6995950"

/tissue_type="neural"

/dev_stage="embryo, stages 14-19"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NICHD XGC Emb6"

/note="Vector: pCMV-SPORE.1; Site 1: NotI; Site 2: EcoRV;

Cloned unidirectionally. Primer: Oligo dt. Average insert

size 2.1 kb. Constructed by Invitrogen. Note: This is a

Xenopus gene Collection (XGC) library."

ORIGIN

Query Match 2.2%; Score 155.4; DB 6; Length 1626;

Best Local Similarity 42.5%; Pred. No. 9.2e-14;
Matches 665; Conservative 0; Mismatches 885; Indels 15; Gaps 2;

```

QY 2709 TGAATTTGGGCAATCTTTTATCAAAATGAAACGAAACGCTTATTTATTAAGTAATC 2768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1615 TTTTATTTTNNATTTTNNATTTTATTAATTTTATTAATTAATTAATTAATTAATTA 1556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2769 TAAATGTCAGATGAATTTATTAACAGACATTTATTAAGAAATTAATGAATTTCAATAG 2828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1555 TTTATTAATTTATTTATTAATTAATTAATTTATTTTATTTTATTTATTTATTTATTT 1496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2829 ATTATCTAAATATGAAATTTTAAAGATATATTAAGTGAATAATCAACATCTTCTCTAT 2888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1495 TTTATATTTTATTTTATTTTATTTATTAATTTTCTTTATTTAAATCTTAATTTTAT 1436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2889 TGTCTTACAGATGATCTATTAATGATGATGCTTAATGCAAGAAATTTAGTTTATTTAT 2948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1435 TTTTATTTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2949 TGTGTAAGATTAATAATGAATAATCTAATTTTAAAGAAATTTACTTAAATAATA 3008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1375 TTTTATTTTATTTTATTTTATTTTATTTATTTTATTTATTTATTTATTTATTTATTT 1316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3009 AAATAGTGAATTTGAGAGTAATGATGATTAATTAATTTAAATGAATGACCGGAATAT 3068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1315 NMAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3069 TTTATTTTAAAGTCTGCGTGAATTTATTTATTCAGAGCAAAAGTATGATTTTAA 3128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1255 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3129 TTAATTTTATGAATTTATTTATTTATTTATTTATTTTATTTTATTTTATTTTATTT 3188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1196 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3189 AAAATGAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1136 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3249 AGTATGTTTGTGAATTAATTTTGAAGATTAATTTGACATTTTACTGCTCCATATTT 3308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1076 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3309 TGAATTTTGAATTAATGATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 3368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1016 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3369 TTAATAAATAGTATCTTTTATTTAGTTTTATTTATTTAGTATATCTGATTTATTTATTT 3428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 956 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3429 ATTCAAAATGGGAAGATATTTTATTTTACAGACACCTTATTAAGACTATGATCTTT 3488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 896 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3489 ATTAACAGCGTCAAAACAAGTGTGGCTTTATGAATCTACAGCTTAATATCAACT 3548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 836 NTTTTATTTTAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3549 ACAATTAATAGTTCAATCCGTAATCTTGCACCTTATTAATAAATAAATGCAACAATTT 3608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 ----TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3609 TTTTCTGTGTCCTGCTTTATTAACGATCTATTTAAGTGAATGAGAAATTTGATTTTA 3668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 730 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3669 TCGCTAGCAATTAATTAATGCTGTTAAGAGATATTAATGCGGAAATTTGCTGTTG 3728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 611
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3729 ATAAATAAGTATATGTAATTTTGTAAATCTACTTATTTATTTTAAATGCAATTTGCTT 3788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 610 ATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 551
QY 3789 TACCATGAATTTTGGCTGTTTATTAATCTAGAGATCAAGTACGAACCTAGATTTATTT 3848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3849 ATTATCAAGAGATTTGATTAATTAAGTATTAAGAAACAATTTTATTTTATTTATTTGA 3908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 NNNNNNANANANANANANANANANANANANANANANANANANANANANANANANANANA 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3909 TCCGAATATTCAGTTACGGGAATTCGCGAAGTCATTCAGGCTATATATCATTTTATTT 3968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 ANNNNNANANANANANANANANANANANANANANANANANANANANANANANANANANA 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3969 TATTAATCAGAAATAGTGGTGTGATTTTATCTGATGTTTCTTTTATTTATGTTATAAA 4028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 TTTTATTTTATTTATTAATTAATTAATTTTATTTTATTTTATTTTATTTATTAATANA 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4029 AAAAGTTATGAGTTATATGCGGAACAGACATTTTATTTTATTTATTCATTTAGCCATA 4088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4089 TTGATATGAACAATAGTTCGATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4149 GGTATTTGGAAATTAATTAATTTTAAAGATATGAGACAAAAATGAATGATTTAAT 4208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4209 TTGATTTATTTGACCAATTTATATGTCGCAAGTTATCTTGATTAATTTGATTAACGAT 4268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4269 TATTTA 4273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 NANNA 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
CG757503 1392 bp DNA linear GSS 24-OCT-2003
LOCUS CG757503
DEFINITION P052-4-C08.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION CG757503
VERSION CG757503.1 GI:37986131
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
            Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1392)
AUTHORS Strinivasan,J., Sizer,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Bunceler,J., Van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@uebingen.mpg.de
            Class: BAC ends.
FEATURES
            location/Qualifiers
            1..1392
            /organism="Pristionchus pacificus"
            /mol_type="genomic DNA"
            /strain="California"
            /db_xref="taxon:54126"
            /clone_id="Ppa EcoRI BAC library"

```

ORIGIN

/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

Query Match 2.2%; Score 155; DB 10; Length 1392;

Best Local Similarity 46.6%; Pred. No. 1.1e-13; Mismatches 602; Conservative 0; Mismatches 683; Indels 8; Gaps 4;

```

2407 AGAATGACACCATTAAGTCTTCAAGAAATACCTGATATATTTTAAATATTCACAGA 2466
103 AGTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
2467 TTTTATGTTGATTTTCAAAAAGATGAGCAAAAATATPAGAAAATGATATATGAAACG 2526
163 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT
2527 AGTTAAATGTTACAGATATATCTTCAATATATATGAGAAAACATATGATATATGATCT
223 TTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
2587 TAGAATTTTATTAAGATATATGACGCTTTGAAATACATTTTACAAAGATTTGTTTAT 2646
283 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT
2647 TGATAGAAATAAAAAATGCTCTAAGAAATATAGATTTGCTCAATTTGGCTTCCCA 2706
343 TTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
2707 CATGATTTTGGCAATTTCTTTTATCAAAATGAAAACGAAACGCTTATTTTATATAGTAA
403 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT
2767 TCTAATATGCCAGATGAACTATTTATACAGACAAATATAGAAAATATGAAATTTTCAA
463 TTTTATATATATATATATATATATATATATATATATATATATATATATATATAT
2827 AGATATCTAATATATGAAAATTTAAGATATATATAGGAAAATATCAACATCTTCCT 2886
523 TAAATTAATATATATATATATATATATATATATATATATATATATATATATATAT
2887 ATGCTCTTACAGATATCTTATGATGAAATTTGCAAAATGAGAAATTTAGTATTTTAA
583 TTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
2947 TTGCTAGAAAGTTTAAATATGAAAATPAAATCTAATTTTAAAGAAATTTTACTTAA
642 TTTTATATATATATATATATATATATATATATATATATATATATATATATATAT
3007 TAAATATAGTGAATTTTGAGAGATATATATATATATATATATATATATATATATAT
702 TTTTATATATATATATATATATATATATATATATATATATATATATATATATAT
3067 ATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
762 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT
3127 TATTAATTTTATGATATATATATATATATATATATATATATATATATATATATATAT
822 ATTTAATATATATATATATATATATATATATATATATATATATATATATATATAT
3187 TAAAAAATGAAATTTTATGTTTATATATATATATATATATATATATATATATATAT
882 TTTAATATATATATATATATATATATATATATATATATATATATATATATATAT
3247 CAAGTATGTTTGTGAATATATATATATATATATATATATATATATATATATATATAT
942 TTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
3305 AATTGATATATGCAATATATATATATATATATATATATATATATATATATATATAT
1002 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT

```

```

3364 AAAAATTTAAAAATATATATATATATATATATATATATATATATATATATATATAT
1062 TTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
3424 ATATATTTCAAAATATGGAAGATATATATATATATATATATATATATATATATAT
1122 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT
3484 ATCTTATACAGCGCGCAAAACAGATGTTGCTTATGACATATCTTACGTTAAATA
1178 TTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
3544 CCACTACATATATATATATATATATATATATATATATATATATATATATATATAT
1238 ATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT
3604 AATTATTTTCTGCTGCTCTTTTATACCATATATATATATATATATATATATATAT
1298 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT
3664 GTTATATGCTAGCAATATATATATATATATATATATATATATATATATATATAT
1358 ATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT

```

RESULT 7

LOCUS

AG386981 1542 bp DNA linear GSS 21-DEC-2004

DEFINITION Mus musculus molossinus DNA, clone:MSHg01-201G10.TJ, genomic survey

ACCESSION AG386981

VERSION AG386981.1 GI:47998186

KEYWORDS GSS.

SOURCE Mus musculus molossinus (Japanese wild mouse)

ORGANISM Mus musculus molossinus

REFERENCE 1 Abe,K., Noguchi,H., Tagawa,K., Yuzuriba,M., Toyoda,A., Kojima,T.,

2 Ewawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriaki,K. and

3 Shiroyoshi,T.

4 Contribution of Asian mouse subspecies Mus musculus molossinus to

5 genomic constitution of strain C57BL/6J, as defined by BAC-end

6 sequence-SNP analysis

7 Genome Res. 14 (12), 2439-2447 (2004)

8 15574823

9 2 (bases 1 to 1542)

10 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

11 Direct Submission

12 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

13 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

14 1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

15 (E-mail:hattori@sc.riken.jp; URL:http://ngp.gsc.riken.go.jp/.

16 Tel:81-45-503-9111. Fax:81-45-503-9170)

17 Clones are derived from the mouse BAC library MSHg01. For BAC

18 library availability, please contact Kuniya Abe (abe@sc.riken.jp).

19 Tsukuba Institute, Bio Resource Center,

20 The Institute of Physical and Chemical Research (RIKEN) 3-1-1

21 Koyadai, Tsukuba, 305-0074 Japan

22 Phone: 81-298-36-9189, fax: 81-298-36-9199

FEATURES

source

1. 1542
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"

ORIGIN

/db_xref="taxon:57486"
 /clone="MSM01-201G10.TJ"
 /sex="male"
 /issue_type="mixture of kidney and spleen"
 /clone_lib="MSM01 Mouse Male BAC library"

Query Match 2.2%; Score 153.6; DB 10; Length 1542;
 Best Local Similarity 46.7%; Pred. No. 1.7e-13;
 Matches 587; Conservative 0; Mismatches 664; Indels 7; Gaps 4;

2604 TGTATCGAGCTTTGGAATCTATTTCACAAAGTTGTTGTTATGATAGATATAAACA 2663
 155 TTATATAAATTTTAAATTAATTAATAAATTTTAAATATATATTTTAAATTTAT 214
 2664 TGTCTAAGAAATAGATTTGGTCTAATGGGTGCTCCACATGATTTGGCAAT 2723
 215 TAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 274
 2724 TCT--TTTATCAATGAAAAAGAAACAGCTTATTTATTAAGTAATCTAAATGCCAGAT 2781
 275 TTTAAATTAATTAATTTTAAATTTTAAATTTTATTTTATATATATATTTTAAATTT 334
 2782 GAACATTTATACAGACATTAATAGAAAAATATGAATTTTCAATAGATTCTAAAT-- 2839
 335 NTTAATTTTAAATTAATTAATTAATTAATTTTATATTTTATTTTATTAATTAATTTTA 394
 2840 ATGGAATTTAAGATATATTAAGCGAAAAATCAACATCTCTCCATATGCTTTACAG 2899
 395 ATATAAATTTTATTAATTAATTAATTTTAAATTTTATTAATTAATTAATTAATTA 454
 2900 ATGATCTATTTGATGAAATGCTAAATGCAAGAAATTTAGGTTTATTTTCTAGAAAGT 2959
 455 ATTTTATTTTATTAATTAATTAATTAATTAATTTTATTTTATTTTATTAATTAATTA 514
 2960 TAAAAATAGAAAAATTAATCTAAATTTTAAAGAAATTAATTAATTAATTAATTAATTA 3019
 515 TTTATATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTA 574
 3020 TTTGTGAGAGTAATGATGTTTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 3079
 575 TTTTATTAATTAATTAATTAATTAATTAATTAATTTTATTAATTTTATTAATTTTATTA 634
 3080 GTACTCTGGTGAATTTTATTTTATTCAGAGCAAAAGATGATTTTATTAATTTTAT 3139
 635 TTAATTTAAATTAATTAATTAATTTTATTAATTAATTAATTTTATTTATTTATTTAT 694
 3140 GAATTAATTTTATTAATTAATTAATTTTGAAGAACTAAGCTAATTAATTAATTAATTA 3199
 695 TTTATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTTAT 754
 3200 TTTATTTGTTTATTAATGAGCTAATTAATGTTTGTTCAGTACTCACAAGTATGTTGT 3259
 755 TTTTATTAATTAATTAATTTTATTAATTTTATTTTATTTTATTTTATTAATTTAT 814
 3260 TGAATTAATTTTGAAGATTAAT--TTGACATTTTACTGCTCCATTAATTTGATTAAT 3317
 815 TTAATTTATTTTATTAATTAATTAATTTTATTAATTTTATTAATTTTAAATTAATTAATTA 874
 3318 GCATTAATGATTAATTAATTTGATTCATTTTAAATTAATTAATTAATTAATTAATTAAT 3377
 875 TATATTAATTTTATTAATTTTATTTTATTAATTAATTTTATTAATTTTATTAATTAAT 934
 3378 AGATCTTTTATTAATGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3437
 935 NTTATTTTATTAATTAATTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTA 994
 3438 GCGAAGATTAATTTTATTAATTTTGAAGCAACCTTAATAGACTAATCTATTAACAGGC 3497
 995 TTTAAATTAATTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTAAT 1054
 3498 GTCAAAACAGGTTGGCTTAATGAACATTAATCTAAGTTAAATACACATTAATTAATTA 3557

Db 1055 TTAATTTATTAATTAATTTTATTAATTAATTAATTAATTAATTTCTATTANTATTAAT 1114
 Qy 3558 GTTTCATTCGCTTAATCTTTCACCTTAATTAATTAATTAATTAATTTTCTTG 3617
 Db 1115 TTTTATTTATTAATTAATTAATTTTAAATTTATTAATTTTATTAATTTTATTT 1174
 Qy 3618 TGTCTGCTTTATTAACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3677
 Db 1175 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATTAATTAATTA 1233
 Qy 3678 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3737
 Db 1234 TTAATTTATTAATTAATTTTATTAATTTTATTTATTTATTTATTAATTAATTAATTA 1293
 Qy 3738 CTAAATGATTAATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3797
 Db 1294 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1353
 Qy 3798 ATTTGCTGTTTATTAATTTAGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 3855
 Db 1354 TTAATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 1411

RESULT 8
 CL509408 1758 bp DNA linear GSS 01-APR-2004
 LOCUS SAIL_811_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone
 DEFINITION SAIL_811_H11.v3, genomic survey sequence.

ACCESSION CL509408
 VERSION CL509408.1 GI:46006728
 KEYWORDS GSS.
 ORGANISM Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1758)
 Sessions,A., Burke,E., Presting,G., Aux,G., McEliver,J., Patton,D.,
 Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
 Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmey,B.,
 Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
 A high-throughput Arabidopsis reverse genetics system
 Plant Cell 14 (12), 2985-2994 (2002)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 PUBMED
 COMMENT
 Contact: Sessions A
 Applied Trait Genetics
 Syngenta Biotechnology Inc.
 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
 Email: allen.sessions@syngenta.com
 ABRC Stock Number CS836276; T-DNA left border flanking sequences of
 Syngenta Arabidopsis Insertion Library (SAIL) lines are available
 through the Arabidopsis Biological Resource Center (ABRC).
 Sequences represent a pool of amplified genomic regions and not
 single contiguous sequences.
 Class: TUNA tagged.

FEATURES
 source
 Location/Qualifiers
 1..1758
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone="SAIL_811_H11.v3"
 /clone_lib="SAIL Collection"
 /note="T-DNA left border sequences were isolated using a
 modified TAIL-PCR strategy"

ORIGIN
 Query Match 2.2%; Score 150.8; DB 10; Length 1758;
 Best Local Similarity 36.3%; Pred. No. 4.6e-13;
 Matches 635; Conservative 0; Mismatches 1084; Indels 31; Gaps 5;

1968 ATAAAAATTTCTGATTAAGAAATGGAACATTAATTAACAAATCGAAGTACTTA 2027

D	27	AAANAATNTNTTATTTAAAAA	AAAAAAAAA	ANTTATTA	TANNNAAAAA	AAAAAAAAA	86				
O	2028	TTTGCACGAGGCCCGCTA	CTTTATGA	ATTCATTA	TCGAAGAAAAA	CAATTA	2087				
D	87	AAANNNNNNNNNNNNNNN	NAANNNNNNNNN	NANATTT	TTTTNNAAAAA	AAAAA	146				
O	2088	TGTTCC	TAGCAAAA	AGTATG	TGAACATG	TAAATG	ATCACTAGAGTTG	147			
D	147	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	206			
O	2148	GAA	GAATTTT	CAAGATAT	TAATTTT	ATTTA	TGAAAA	ATGATGATTTG	2207		
D	207	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	266		
O	2208	AAATTA	TGAA	CTT	TAAGCA	ACTTA	CTTACATCA	ATTAATTTT	TGCGAA	2267	
D	267	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	326	
O	2268	GATTA	AAACCA	ATAGTT	GAAAAATTTA	TGAG	GATCAAGAA	ATGATTA	TAAAAAGAT	2327	
D	327	NNNN	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	383	
O	2328	GCAT	TTTGA	TAA	TGCGCTTA	CA	TAA	TTTTCT	CAGATTTTA	CTGAGAGGAT	2387
D	384	AAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	443	
O	2388	ATTAT	CACTCT	CTCAGAG	ATG	CACACATTA	GTCTCT	CAGAA	TACCTG	TAAAT	2447
D	444	AAAA	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	503
O	2448	ATTT	AAAT	ATTTCT	CAGATTTA	TATGTTGA	ATTTCA	AAAG	ATGAGCA	AAATAT	2507
D	504	AAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	563
O	2508	AAAT	TAGAT	TATGA	CAGAGTTA	ATGTTCA	GTATTTCT	CTATAT	TATCA	GAAAA	2567
D	564	AAAA	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	623
O	2568	CTAT	TGAT	TAAGT	ACTGTT	TGAATTTT	ATTA	GAGATG	TATCGAGCTTT	GATTA	2627
D	624	AAAA	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	675
O	2628	TACA	AGAT	TGTTT	TATGAT	ATAAAA	CATG	GTCTA	GAATTA	GATTTG	2687
D	676	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	735
O	2688	TAT	TGGG	CTT	CGCTT	CA	CATGATTTT	TG	CGCAATCTTT	TATCAATG	2747
D	736	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	795
O	2748	AGCT	TATTTATTTA	GTATCT	AAATG	CCAGAT	CACT--	ATTAT	ACAGACAT	TATA	2805
D	796	ANNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	NNNN	855
O	2806	GAAA	AT	TGAATTTT	CAATAG	ATATCT	AAATAG	AAATTTA	AGATAT	ATAAGT	2865
D	856	AAT	ATAA	TATATAA	TATATA	TATATA	TATATA	TATATA	TATATA	TATATA	915
O	2866	AAAA	ATCA	CACTCT	CTCTAT	GTGCTTTA	CA	CGATGATCT	TATGAT	GAAATG	2925
D	916	AAAT	TAAT	TNNAAAA	ANTTAT	TAAATTA	ATTAATTT	TATTAATTT	TATTAATTT	TATTAAT	975
O	2926	GCA	GAAATTA	TAGGTTT	TTTATTTG	CTAGAA	GTAAAGTTAA	ATAGAA	ATTAATCT	AAATTT	2985
D	976	AAT	TAAT	TATTAATTT	TATTAATTT	TATTAATTT	TATTAATTT	TATTAATTT	TATTAATTT	TATTAATTT	1035
O	2986	AAAG	AAATTA	TATCT	TAATAA	ATTAATAG	TGATTTG	TGAAGT	TAATGAT	TGTTTAAT	3045
D	1036	AAAA	ATTA	TATAA	ANTTATTA	TATAA	TATATA	TATATAA	-----	-TAT	1082
O	3046	TAT	TAAAT	TATGAC	CCGGA	TATTTT	TATTTT	TATGAT	CTCTG	TGATTA	3105
D	1083	TAT	TATAT	TATATTA	TATATATTTT	TATATTA	TATATTA	TATATTA	TATATTA	TATATTA	1142

QY	3106	CAGGCAAAAGCTATGTTATTTTATTAATTTTATTAAGAACTTAACTTAACTTATTCGATPATAAT	3165
QY	3106	CAGGCAAAAGCTATGTTATTTTATTAATTTTATTAAGAACTTAACTTAACTTATTCGATPATAAT	3165
Db	1143	ATANATATATATTAATAAATAATTAATAAATAATTTAAATTTAAATTTAAATTTAAATAAT	1202
QY	3166	TTTTGAAAACTACACCTATATTTAAATAATGAAATTTTATGTTTATTTATGATCTATAT	3225
Db	1203	TATTTAAATATATTTAAATATTTAAATATTTAAATATTTAAATATTTAAATATTTAAATAT	1262
QY	3226	TATGTTTGTGTTCACTAGTCAAGATATGTTTGTGAATTAATTTTGAAGAATTTATTTG	3285
Db	1263	TNTTATATTAATATTTAAATTTAAATAATATAAATNTAANTAAATTTANNNNAATAATAT	1322
QY	3286	CAGATTTTACGCTCCCATTAATTTGGATTTGCAATPAATGTATTAATTAATTTGATTCAT	3345
Db	1323	TAAATATATATNNNTATTTATTTATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAT	1382
QY	3346	TTATTAATATTTGATTTAAATAAATAAATAAATAAATAGTATCTTTTATGTTTATGTTTAT	3405
Db	1383	TNNAATATTTATTTATTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA	1442
QY	3406	TAGGTATATCTGCATTTGTATTTATTTCAAAATGGGAAAGATTTGTATTTTGAACAGAC	3465
Db	1443	AAAAATTTAAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1502
QY	3466	ACCTTATAGGCTGACATCTTTTAAACAGCGTCA- - -AACAGGCTGTGTCCT	3520
Db	1503	AATTAATTAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	1562
QY	3521	TATGACATCTGCTGCTTAATTAACACCTACATTTATGTTCAATTCGTTATCTTTGC	3580
Db	1563	TTATTTATATATTTTAAANNANTTAATATATATTTATTTATTTATTTATTTATTTATTT	1622
QY	3581	ACTTATTAATAAATAAATGCAACAATTTTTTCTGTGCTCTTTATTAACGATCTA	3640
Db	1623	TTAAAAATATATNTANNNTNTNTTTTATATTAATAATTAATTAATTAATTAATTAATA	1682
QY	3641	TTTAAAGGATCGGAATTTGGTACTTTATGCTGACATATTTATTTATGCTGTTATG	3700
Db	1683	TATNTTAAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	1742
QY	3701	GAGATATATTA 3710	
Db	1743	AAAAATATA 1752	
RESULT 9			
CLS09408/c	1758 bp	DNA	linear
LOCUS	SATL_811_H11.v3	SATL Collection	Arabidopsis thaliana genomic clone
DEFINITION	SATL_811_H11.v3	genomic survey sequence.	
ACCESSION	CLS09408		
VERSION	CLS09408.1	GI:46006728	
KEYWORDS	GSS.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryote, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, eudicotyledons, core eudicotyledons, rosids, eutroside II, Brassicales, Brassicaceae, Arabidopsis. 1 (bases 1 to 1758)		
AUTHORS	Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Paton, D., Dietrich, B., Ho, P., Bacmadel, J., Ko, C., Clarke, J. D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M. and Gelf, S. A.		
TITLE	A high-throughput Arabidopsis reverse genetics system		
JOURNAL	Plant Cell 14 (12), 2985-2994 (2002)		
PUBMED	12468722		
COMMENT	Contact: Sessions A Applied Trait Genetics Syngenta Biotechnology Inc. 304 Cornwallis Rd., Research Triangle Park, NC 27709, USA Email: allen.sessions@syngenta.com ABRC Stock Number CS836276; T-DNA left border flanking sequences of		

Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number CS836276; T-DNA left border flanking sequences of

RESULT 10				
AJ925855/c				
LOCUS	AJ925855	1238 bp	mRNA	linear
DEFINITION	Theileria annulata merozoite Theileria annulata cDNA clone			

ACCESSION lam009a05_g1k, mRNA sequence.
 VERSION AJ925855
 KEYWORDS EST.
 SOURCE Theileria annulata
 ORGANISM Eukaryote; Alveolata; Apicomplexa; Piroplasmida; Theileridae; Theileria.
 REFERENCE 1 (bases 1 to 1238)
 Authors Pain, A., Renaud, H., Berriman, M., Murphy, L., Yeats, C. A., Weir, W., Keshornou, A., Ajelet, W., Bishop, R., Bouchier, C., Cochet, M., Coulson, R. M. R., Cronin, A., de Villiers, E., Fraser, A., Foster, N., Gardner, M., Gobie, A., Griffiths-Jones, S., Harris, D. E., Katzer, F., Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F., Nene, V., O'Neill, S., Price, C., Quail, M. A., Rabinowitch, E., Rawlings, N. D., Rutter, S., Saunders, D., Seeger, K., Shah, T., Squares, R., Squares, S., Tivey, A., Walker, A. R., Woodward, J., Dobbelaere, D. A. E., Langsley, G., Rajandream, M. A., McKeever, D., Shiels, B., Tait, A., Barrell, B. and Hall, N.
 The genome of the host-cell transforming parasite *Theileria annulata* and a comparison with *T. parva*
 Unpublished (2005)
 Contact: Pain, A.
 The Pathogen Sequencing Unit
 The Wellcome Trust Sanger Institute
 Genome Campus, CB10 1SR, UNITED KINGDOM
 Merozoite CDNA library; Frank Katzer and Brian Shiels, Division of Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
 FEATURES
 source
 1..1238
 /organism="Theileria annulata"
 /mol_type="mRNA"
 /isolate="Ankara (clone D7)"
 /db_xref="taxon:5874"
 /clone="lam009a05_g1k"
 /dev_stage="merozoite"
 /lab_host="Bos taurus (cow)"
 /clone_lib="Theileria annulata merozoite"
 /note="Country: Turkey; Ankara"
 ORIGIN
 Query Match 2.1%; Score 149; DB 1; Length 1238;
 Best Local Similarity 47.2%; Pred. No. 8.9e-11;
 Matches 582; Conservative 0; Mismatches 637; Indels 15; Gaps 4;
 QY 2810 AATATGAAATTTTCAATAGATTAATCTAATATGAAATTTAAGATATATAAGTGAAA 2869
 DB 1238 ATATATTTAATTAATATATATATTAATATTAATATTAATTAATTAATTAATAT 1179
 QY 2870 AATCAACATCTTCCTCATGTCTTTACAGATGATTCATGATGAATGCTAAATGCAA 2929
 DB 1178 AATATATTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1119
 QY 2930 GAATTTAGGTTTTTTTATGCTAGAAAGTTAAATAAGAAATTAATCTAATTTAAG 2989
 DB 1118 ATTTTATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1059
 QY 2990 AATATTTACTAAATAAATAAATAGTTGATTTTGAGAGATGATGATGTTAAATTAAT 3049
 DB 1058 AATATTTAATTAATAAATTAATTTTATTTATTAATTAATTAATAATTAATTAATTA 999
 QY 3050 TAAATATGACCGGAATATTTTATTTTAACTCTCGTGATTTATTTTATCCAGA 3109
 DB 998 TTATTTATTTTAAATTAATTAATTTATTTTAAATTAATTAATTAATTAATTAATTT 939
 QY 3110 GCAAAAGATGATTTTATTTTATTTTATGAAATTTATTTTATTCATATAAATTTT 3169
 DB 938 ATTAATTTATTAATA-TATTTATTAATTTTATTTTATTTTATTTATTAATTAATTTT 880
 QY 3170 GAAATCTAAGCTAATTAATTAATAAATGAAATTTTATG-----TTTATTAATGCTCAT 3223
 DB 879 TTTATATTTTATTTATTTATTTATTTATTAATTAATTTTAAATAATTAATTTTATTTAT 820

QY 3224 ATATATGTTTGTTCAGTACAGTATGTTGTTGAATAAATTTGAAAGATTAAT 3283
 DB 819 ATTAATTTTATTAATTTTATTTTATTTTATTTTATTTTATTAATAATTTTATTAATTA 760
 QY 3284 TGCAGATTTTACGCTCCCAATTTGGATTTATGCAATAATGATTAATTTGATTC 3343
 DB 759 TATTTTATTTTATTTAATAATTAATTAATTTTATTTTATTTAATAAATTTAATTAAT 700
 QY 3344 ATTTATTAATATGATTTAATAAATAATTAATAATTAATGATCTTTTATGTTTATGATTT 3403
 DB 699 TTTATTTATTTATTAATTTATTAATTAATTAATTAATTAATTTTATTTATTTATTTTAT 640
 QY 3404 ATTAGATATATCTGATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3463
 DB 639 TTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 580
 QY 3464 ACACCTTAATAGACATGATCTTATTAACAGCGGTCAAAACAAGTTGGTGGCTTAT 3523
 DB 579 TTTATTTTATTTTATTAATTAATTAATTAATTAATTTTATTTTATTTTATTTTATTTAT 520
 QY 3524 GAACATAT-----CCATAGTTAAATACACATGATTAATGATTCATTCGTTATGCTT 3578
 DB 519 TAAATTTATTAATTAATTAATTAATTAATTAATTAATTTTATTTATTTTATTTATTAAT 460
 QY 3579 GCACTTTAATAAATAAATAATGCAACAATTTTCTGTGCTGCTTTATACGATC 3638
 DB 459 TTTATTTTATTAATTAATTAATTTATTTATTAATTTATTTATTTATTTATTTATTTAT 400
 QY 3639 TATTTAATGATCGAATTTGATTTATTCGTACAAATTAATTAATTAATTAATTTGCTGTTA 3698
 DB 399 TTTTATTAATTTATTAATTAATTAATTAATTTTATTTTATTTAATTAATTAATTAATTA 343
 QY 3699 TCGAATATATAGTGTGAAAATTTGCTGTGATTAATAAAGCTAATAGTAATTTGTATA 3758
 DB 342 TAAATTTTATTAATTAATTAATTAATTAATTTATTTATTTTATTTATTTATTTATTTAT 283
 QY 3759 CTACTTTATTTTAAATTAATGCAATTTGCTTACATGAATTTTGGCTTTATTAATTTCT 3818
 DB 282 TTTATTTATTTATTTATTTATTTATTAATAATTAATTAATTTTATTAATTTATTTATTT 223
 QY 3819 AGAATCAAGTAAAGAGCTAGATTTATTTATTAATCAAGAAATGATTAAGTAAAGTAA 3878
 DB 222 AATTAATTAATTTATTAATTAATTTATTTATTTATTTATTTATTAATTAATTTATTTAT 163
 QY 3879 GAAACATATTTTATTTGATATGAAATATCCGAATTTCAATTTAGTACGGAATCTGGCTC 3938
 DB 162 TTTATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA 103
 QY 3939 GGAAGCATTCAGGCTATATATCAATTTTATTAATCAGAAATGATTTGGGTGATTTTA 3998
 DB 102 ATTTTATTTTATTTTATTTATTTATTAATTAATTAATTAATTAATTAATTTATTTATTT 43
 QY 3999 CTGATGTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4032
 DB 42 CATATTAATTAATTTATTTATTTTATTTAATAAATAAATAAATAAATAAATAAATAAATA 9
 RESULT 11
 AG320553 1780 bp DNA linear GSS 18-DEC-2004
 AG320553
 LOCUS
 DEFINITION Mus musculus molossinus DNA, clone:MSW01-106h14.TJ, genomic survey
 ACCESSION AG320553
 VERSION AG320553.1 GI:47893510
 KEYWORDS
 SOURCE GSS.
 ORGANISM Mus musculus molossinus (Japanese wild mouse)
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 Abe, K., Noguchi, H., Tagawa, K., Yuzuriba, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hatori, M., Sakaki, Y., Moriaki, K. and

RESULT 12
AG350139/c
LOCUS AG350139 1489 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSM901-146K14.TU, genomic survey
sequence.
ACCESSION AG350139
VERSION AG350139.1 GI:47923449
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaiki,K. and
Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1489)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Shuhiro-chou,Tsukumi-Ku, Yokohama, Kanagawa, 230-0045, Japan
(e-mail:hattori@gsc.riken.jp, URL:http://bgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSM901. For BAC
library availability, please contact Kunihisa Abe (abe@rc.riken.jp).
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rc.riken.jp
PRIMERS
Sequencing : TU
LIBRARY : PBACe3.6
Vector : EcoRI
R.Site 1 : EcoRI
R.Site 2 : EcoRI
FEATURES
source Location/Qualifiers
1..1489
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSM901-146K14.TU"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_lib="MSM901 Mouse Male BAC Library"
ORIGIN
Query Match 2.1%; Score 146.6; DB 10; Length 1489;
Best Local Similarity 48.3%; Pred. No. 2e-12;
Matches 646; Conservative 0; Mismatches 671; Indels 21; Gaps 8;
QY 2052 TTATGAATTCATTCCTCAAGAGAAAAACAATATGTTCCCTAGACAAAGAAAGATG 2111
DB 1455 TTTAT 1396
QY 2112 GTGACACATGTAATGATCATCAAGTAGAGTTGTGAAGAGATTTTCAAGATATATA 2171
DB 1395 ATTAT 1336
QY 2172 TTTTAT-TTATAGAAAATATAGATGATTTGTTGAAAAAATATATGAAGTTCTTAACCA 2230
DB 1335 TAGTA 1276

QY 2231 ACTAATCTTACATCAATCAATTAATTTTTTTTGTGAAAGATTTAAACAAATAGTTGAAAA 2290
DB 1275 NATTTAATATTAATTAATTAATATATATATTTTTAGATTATATATAGTAATTTATATATA 1216
QY 2291 TTTATAGAGATCAAGAAAATGAATTAATTAAGAAAAGATGCAATTTTGATATGCTTATCA 2350
DB 1215 TTTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1157
QY 2351 TAAATTTCTCAGATTTTACGTGGAGAGATGATAGATATATATCATCTCTCCAGAGAA 2410
DB 1156 TTATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1097
QY 2411 TGACACACATTAAGTTCCCTGAGAAATACCTGATATATATTTTAAATATCTCAGAAATTA 2470
DB 1096 TTAATTTAAATTTTAAAGTTTATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 1037
QY 2471 TAGCTGAATTTACAAAGATGAGCAAAATATTAAGAAAATAGATATATGAAAGAGTT 2530
DB 1036 ATTAAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAAGANTTAATTAATTAATTA 977
QY 2531 AAATGTTACAGATTTATTTCCATATATATCAGAAAACCTATGATATATGTTAGTCTGTTAGA 2590
DB 976 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 917
QY 2591 ATTTTATTAAGAAATGATTCGAGCTTTTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2650
DB 916 ATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 857
QY 2651 AGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2710
DB 856 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 797
QY 2711 ATTTTGCGCAATTTCTTTTCAATGAAACGAAACGCTATTTATTAATTAATTAATTAATTA 2770
DB 796 ATTAATTAATTAATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 737
QY 2771 AATGTCAGATGAACTTTTATACAGCAATTAATTAAGAAAATTAATTTTCAATAGAT 2830
DB 736 TTATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 677
QY 2831 TATCTAAATATGAAATTTAAGATATATTAAGTGAAGAAAATCAACATCTTCTCTATG 2890
DB 676 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 621
QY 2891 TCTTTACAGATGATTCATTTGATGAAATGCTAAATGCAAGAAATTTAGTTTTTATTTG 2950
DB 620 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 566
QY 2951 CTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3010
DB 565 TAAATTAATTAATTAATTTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 507
QY 3011 ATAGTTGATTTGTGAGATGATATGATTTTAAATTTTAAATTAATTAATTAATTAATTAATTT 3070
DB 506 TTATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 447
QY 3071 TATTTTAAATTAATTAATTAATTTTATTTTATTTTATTTTCCAGCAAAAGATGATTTTATTT 3130
DB 446 TAAATTAATTAATTAATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTA 390
QY 3131 AATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 3190
DB 389 ATTATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 330
QY 3191 AAATGAATTTTATGTTTATTAATTAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAAT 3250
DB 329 AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 275
QY 3251 TATGTTTGAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 3310
DB 274 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 216
QY 3311 GATTAATTAATTAATTAATTAATTTGATTCATTAATTAATTAATTAATTAATTAATTAATTAAT 3370

Db 215 TATATTTTATATTAATTTATATATATTTTAAAAATTTATTAATTAATT 156
QY 3371 AAAAAATGATCTTTT 3388
Db 155 AAAAAATTAATTAATTAAT 138

RESULT 13
DN685273/c 1359 bp mRNA linear EST 30-MAR-2005
LOCUS CGX40-G12 5', mRNA sequence.
DEFINITION CGX40-G12 5', mRNA sequence.
ACCESSION DN685273
VERSION DN685273.1 GI:62033458
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1359)
Kingley,D.M., Pelchel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Grimwood, Jane
Stanford Human Genome Center
Stanford University School of Medicine
975 S California Ave, Palo Alto, CA 94304, USA
Tel: 650 320 5917
Fax: 650 320 5801
Email: jane@hgc.stanford.edu
Plate: 40
High quality sequence start: 18
High quality sequence stop: 102.
Location/Qualifiers
1..1359
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Conner Creek sticklebacks, WA USA"
/db_xref="taxon:69293"
/clone="CGX40-G12"
/sex="mixed male and female"
/tissue_type="eyes"
/dev_stage="adult"
/lab_host="DH10B (T1 phage resistant)"
/clone_idb="SHGC-CGX"
/note="Vector: Express 1: Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand primer: 5'-GACTAGTCTAGATCGAGCGCGCC(T)25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC' (where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at:
http://www.openbiosystems.com/cdna_library_construction_fa9.php#8 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems:
http://www.openbiosystems.com/stickleback"

ORIGIN
Query Match 2.1%; Score 144.8; DB 8; Length 1359;
Best Local Similarity 47.5%; Pred. No. 3.8e-12;

Matches 536; Conservative 0; Mismatches 565; Indels 27; Gaps 6;
QY 2156 TTACAGATTAATTAATTTTATTTATTAAGAAAAATATGATGATTGTTGAAAAATTAAT 2215
Db 1238 TAATANAATANNTTAAAAATTTNNNNAAAAAATAATTAATTAATTAATTTT 1179
QY 2216 GAAGTTCTAAGCAACTAATTTACATCAATTAATTTTGTGAAAGATTAA 2275
Db 1178 TAAAAAATTTTTTTTTTATTAATTAATTAATTAATTTTGTGAAAGATTAA 1119
QY 2276 CAATAGTTGAAAAATTTAATGAGATCAAGAAATGAATTAATTAATTAATTAAT 2335
Db 1118 AATANAATTTTAAATTAATTAATTAATTTNNNNTTTA-AATTAATTTTAAATTTT 1060
QY 2336 GATTAATGCTTATCATTAATTTTTCAGATTTTACTGAGAGGATACGATTTATCAT 2395
Db 1059 AATNTAATAAANNTTTTNTTTTAAAAAATATTAATTAATTAATTAATTTAAAT 1000
QY 2396 CTTCCTCAGAGAAATGCACACATTAATTCCTGAGATTAATCCTGATTAATTTAA 2455
Db 999 TTATTAATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTTATTA 940
QY 2456 TATTCAGGATTTATATGTTGAATTTACAAAAGATGAGCAAAATATTAAGAAAAATAG 2515
Db 939 AATTANAATTAATAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 880
QY 2516 ATATATGAACGAGTTAAAT-GTTACAGATTATTTCTAATATATCAAAAAATCTATGA 2574
Db 879 TTTTATTAATTAATTAATAAATAATTTTATTAATTAATTAATTAATTAATTAATTA 820
QY 2575 TAATGATCGTTTGAATTTTATTAAGATATGATCGAGTTTGAATTAATTAACAA 2634
Db 819 TAAANAATTTTAAAAAANAATAATTAATTA--ATTATTAATTAATTAATTAATTA 763
QY 2635 ATTTGTTTATTAATGATGAATTAATAAACAATGCTTAAGATTAAGATTGTTCTAAAT 2694
Db 762 ATATTAATTAATAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 703
QY 2695 GTTTCGCTTCACATGATTTTGTGGCAATTTCTTATCAAAAGAAAAGCAACGCTTAT 2754
Db 702 TTTT---TTANNAATATATTAATTAATTAATTTATTAATTTTAAAAATTAATTAAT 646
QY 2755 TTATTTAAGTAATCTAATGTCAGATGAATTTATTAACAGCAATTAATTAAGAAATAT 2814
Db 645 AATTTTAAAAAATTTTAATTAATAATTAATAAATAATTAATTAATTAATTAATTAAT 586
QY 2815 GAATTTCAATGATTAATTAATAATTAAGAAATTTAAGATTAATAAGTGAATAATCA 2874
Db 585 ATATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 526
QY 2875 ACATCTTCCTATTTGCTTTACAGATGATTTGATGATTAATTTATTCAGAGCAAT 2934
Db 525 ATATTAATTAATAAATAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 466
QY 2935 TTAGGTTTTTATTTGCTAAGAAATTAATAATTAATTAATTAATTTAATTAAGAAAT 2994
Db 465 ATATTTTAATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 406
QY 2995 ATTAATTAATAAATAATTAATTAATTTGTGAGATGATGATTTAATTAATTTAAAT 3054
Db 405 AATTAATAAATTTTAA-----TTAATAATTAATAAATTAATTAATTAATTAATTA 359
QY 3055 ATGACCCGGAATTTTATTTTAAAGTATCTGATGATTAATTTATTCAGAGCAAA 3114
Db 358 AANAATTAANAATAATTTATTAATTTTAAATTAATTAATTTTAAATTAANAATAAT 299
QY 3115 AGATGATTTTATTAATTTTATTAATTAATTAATTAATTTATTAATTAATTTTGA 3174
Db 298 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 245
QY 3175 CTAAAGCTAATTAATAAAGAAATTTATGTTTATTAATGCTATATTAATTAATGTTT 3234
Db 244 ATTTAATTAANAATTAATTAATCAAAATTAATTAATTAATTAATTAATTAATTAATTA 185

Qy 3235 TTTCAGTAGTCACCAAGATGTTGTTGAAATAAAATTTGAAAGATTAT 32822

Db 184 TATANTNAATTAATATATATATAATAATANAATAAANAATAAAATAT 137

RESULT 14

LOCUS	1489 bp	DNA	linear	GSS 18-DEC-2004
DEFINITION	Mus musculus molossinus DNA, clone:MSMG01-146K14.T17, genomic survey sequence.			
ACCESSION				

ACCESSION	AG350139	
VERSION	AG350139.1	GI:47923449
KEYWORDS		
CCS		

SOURCE Mus musculus molossinus (Japanese wild mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE AUTHORS

TITLE	Author
Contribution of Asian mouse subspecies <i>Mus musculus molossinus</i> to the spread of <i>Leishmania</i> in the USSR	Shiroishi, T.

genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
Journal Genome Res. 14 (12), 2439-2447 (2004)

JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1489)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

JOURNAL
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenho-Chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: hattori@gsc.riken.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT
Clones are derived from the mouse BAC library MSNg01. For BAC

library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp

```
Sequencing : T7
LIBRARY
Vector      : pBACe3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI.
```

FEATURES	Location/Qualifiers
source	1. .1489

```

/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSM91-146X14.1J"
/sex="male"
/tissue_type="mixture of kidney and spleen"
clone_lib="MSM91 Mouse Male BAC Library"

```

ORIGIN

Query Match	2.1%;	Score 144.4;	DB 10;	Length 1489;
Best Local Similarity	46.7%;	Pred. No. 4.4e-12;		
Matches 643;	Conservative	0;	Mismatches 699;	Indels 34; Gaps 5;

Qy 2115 AACATGTAAATGATCATCAGTAGAGTTGTGAAGAAGAAATTTTCAAGATA---ATAATA 2171

Db 111 AATATGTGTGTGACCACTGCTGTGTATATATTTTATAATTTTAAATATATATTAATA 170

Qy 21 2 TTTTATTTATAGAAAATATAGATCATTTGTTGAAAAAATTAATGAAGTTCTTAAGCAA 223
Db 171 ATATTTTAAATAAATTATATATATATAAATTATATATATAAATATATTTTATTAATTAATA 230

Qy	2232	TACCTTACACCAAAATATATTTTGTGCAAGTTTAAACAAATACGTGAAAAT	2291
Db	231	ATTAATTTTTRATATTTATTTTATATATTTTAAATATTAATTTTAT	290
Qy	2292	TTAATGAGATCAAGAAAATGATATATATTAAGATGATATTTGATATGCGTTATCAT	2351
Db	291	ATAATATATTTATTTATATTTATTAATATTAATTTTAAATATTTTATTTATAT	350

QY 2352 AATTTTCTCAGATTTCCTGGAGAGGAGATACAGATATTATCATCTTCTCTCA-GAGAGA 2411L

Db 351 TATATTTTAAAAAAATTTAAATTTAAATTTATATATTTATTTATTTTAAATTTTTTT 410

[illegible][illegible]

DY 2531 AATGTTACAGATTATTTCCCTAATAATACAGAAAAAACTATTGTAAATGACTGGTTAGA 2590
| | | | |
| | | | |
DB 531 ATTTTAAATTAATATATATAATTTTATTTTATATTTTAAATTAATATTTATTTTAAA 590
| | | | |

Oy 2591 ATTTATTAAGATGTACGAGCTTTCATACTATTACAAGAATTGGTTAATGCAT 2650

D8 ||||| | ||||| ||||| ||||| |||||
591 ATTAAATTTAAAATTTAAAAATTTAATTTAATTTAATTTAATTTAATTTAATTTT 650

2651 AGATAAAGACATGCTAGATAAGATTGGTCTAATGGGTTTCGCTCCACATG 2710

[illegible]

2748 AGCTATTATTAAAGTATCTAAATGTCAGATGAACATATTATACAGACAATTATAGA 2807

2808 A A A A T A T G A T T T C A A T A G A T T A T C T A A A T A T G A A A T T T A G A T A T A T A A A G T G A A 2867

2868 AAATCAATCTTCTCTATGTCTT-----TACAGATGATTCTATTGATGATGCTTA 2923

2924 ATGCAAGAAATTAGGTTTTTTATTGCTAGAAAGTTAAAAATGAAAAATAAATCTAAT 2983

2984 TTAAGGAATTACTTAAAAATAAAATAGTTGATTGTTGAGAGATATGTAATGTTAA 3043

3044 ATTATTAAATATGACCCGGAATATTATTAACTACTCTGGTGATTATTTTAT 3103

3104 TCCAGAGCAAAAGTATGATTTTATTAAATTTTGAAGATTAAATTTATTTTCATATATAA 3163

3164 ATTTTGAACCTAGCTAATAATAAAAAATGAATTTTATGTTTATTATGCTAT 3223

3224 ATTATGTTTGTTCAGTAGTCACAGATGTTGTGAAATAATTGAAAGATTATT 3283

DG
ACGATTAATTATTTTAACTAAGTGGTGTGTTTGTTTGA 3340

OY 3384 TGCAGATTTTACTGTCCTCCATAAT---TTGCAATTATGCCAATAATGTAATTAAATTGTA 3340

Qy	3447	ATTGATTTT	3457
Db	1500	ATTTAATTAT	1510

Search completed: December 25, 2005, 16:11:32
Job time : 16557 secs

THIS PAGE BLANK (USPTO)

QY	133	TTTTTATATCTCGATATGCCAGTTGAATTTAGATATAGAGTAATCTGATAGAGTTTGA	192
Db	558	TTTATCTTTTCGATTTTTTACAGAACTTTTGGAGTCGTGCTATCTTGAAGAGCTTTAA	617
QY	193	AAACATTTAATCATAGTATTAATATTTGGCAATTTTTCTTACGGCGAGATCAATTTTG	252
Db	618	TGATTTGAATACAGCTTTTACTATATTTTGCATATCAAGTTCAATTTATTTATTTCT	677
QY	253	AGAAATATTTGCACTTTCAGAGACGTGTGCGGTATTTTCATATTAATTAACCTTG	312
Db	678	AAACCTTTTACACAGACGAGCTTTCCTTTTATCTTTTATGTGATGAATTCGAT	737
QY	313	TGATATACCTTATTACGTAAATTTATTAAGCAGTTTAAGATAGCTTTCTATTTTCC	372
Db	738	TATTTATCTATTTGAATTCATTTTTTAAATTTATTCGAAATATTTTACGCTA	797
QY	373	TCTATCAAAAAAAGACGATCTTAATTTACAGCGCTGACATGCGGAATATATG	432
Db	798	CACGAGATACCAAGTTGTTTGTATACGAATTAAGATTTCTTTATCCAAAATG	857
QY	433	TATTTGAATCACTATAACAAATTCGTTTTGTCATTTGATGTTTATAGTA	492
Db	858	GGAAATTAAT---ACGACCATTAATATATCGCTGCTGATCTTGATTTCC	914
QY	493	AAATGATTAATTAATTTATATCATTCACGCTCTATATCTGTGGAAGAACT	552
Db	915	ATTTGTAATGATTTGAAACATATCTGTTAAGATTAATTAACAAAGATCTT	974
QY	553	TTTCAACAAGGAGATGGTCGACACGCTTTATTAATCTACCAAGTGA	612
Db	975	AGTTAACCTGCTTAACCTTGATCAAGCTTTTATTAATACATACCATTA	1034
QY	613	TAAACCAATTC-----GTTTCAGATTTTGAAGTTGATTAAGTAAG	663
Db	1035	AATACCAATTAACAAGTATTAATTAATGACATTTGAAGCAATGGAAT	1094
QY	664	ATATTAAATTCATTCGTTTACTGCGTTGAAGAAACAAAAATTCAC	723
Db	1095	ATGTAGAGCACTTATGCTTGTATTAATTAAGGAAACCGAATCA	1154
QY	724	ATAGCATTTGTAATTTTCCAAATTTTAAATTAAGCTATATCATATG	783
Db	1155	ATAGGTTATTAATCATATTTATGAATTTCTATTAATTAATGTA	1214
QY	784	TTTTGGATTTACTCGAGCGGTAGTCGGTTATTAATTTGTTAGTT	843
Db	1215	TTTTGGATTTCAATGGGTCTATTAATAGTTTCTCATATGTGCAAT	1274
QY	844	TAGTTCCAAATTTTGTAGAGATGTGACCGGCTATTTTGTCTCAG	903
Db	1275	TAGTTCCGCAAAATCAGAAATGATGGTGACCGGCTATCTTCTCA	1334
QY	904	AGAAATGAGCATTTTATTAATTTCAATTTCAATTTGCATATGAT	963
Db	1335	GTAATGTAGATTTTATTAATTTCTAATTAATTAAGATCAATG	1394
QY	964	GCAAAAAGACTTGTCCAGCAAAAACAGATCAAGGGTGTATGTTT	1023
Db	1395	TTTAGAAAGATTTTATTTAGTTCAATCA-----AATGACGGG	1449
QY	1024	AAACGATCTAGAAATTTACTCAATTTGACATTTTATGCAAAA	1083
Db	1450	CGATGATCCTAGAATTAATAAATGAGAAATTTATTTGCG-AAA	1508
QY	1084	TACCAAGTTTATTAATGTTTATTTGAGATATGAGTCTACTTGT	1143
Db	1509	TGCTCAATTTCTAATATGTTTAAAGGATATGATTTATAGAA	1568
QY	1144	CAGTTGATTAATTTGAAATATATATCTGTCGCAAAAGAGAT	1203
Db	1569	CAATTAATGATATGAAAGATATATTTCAACGCAAGGAGAC	1628
QY	1204	GGATTAACAGTCTCTGCGAGGTTAGTGTCGTATTAATATCA	1263

Db	1629	GAATCAGCTGGTTGTGGCAAAATATCTGGTAGAAATATATTAATCTGATTTTGGATGAATCG	1688
Qy	1264	TTCCGTTTGACCTTAGACATACATTAATGAATGAATCTATCTGTCAGATATATTAATTTAT	1323
Db	1669	TAAAGTTAGATGTTCAATATATCAATGAATAGCTATTTGGTCAGATATTAAGATTAATTC	1748
Qy	1324	TAAAGACAGTAAAGTTGTATTGTTGAGAGAGGAAGTAATGAAGTA-----	1372
Db	1749	TCCTTAACGCTAAGAGTGGTTTACTCCGGACAGAGCTAAGTAAAGTAAGGTTTGAAG	1808
Qy	1373	-----TATGAAGTTTGTGTGGCCGCTTCCAGGGGGACATTAAGTCACTTGATTT	1426
Db	1809	GAATATATAGTAAATTTGTCTGTTGGTTAAGTGGTGTCACCTTAGCACATTTGAACCT	1868
Qy	1427	GTTAAAAACCGTTTGGAAAGAAAGAACGTTTTTGGGTAACTTTGATTAAGAGATGC	1486
Db	1869	TTTGAAGCCCATTTGGGAAAAAGAAAGATAGTTTGGGTAACTTTTGAATAAAGATGC	1928
Qy	1487	AAGAAGCTTTTGAAGATGAATAAATGTATCCATGTTACTTTCCAAATATGCATCT	1546
Db	1929	TAGAGATATCTAAGAAAGAGATGTGTATATCATTTGCTTTTCCAAACCGTATATG	1988
Qy	1547	CATTAATTTAGTAAAAATCTTTCTAGCTTCAAAATTTTAGCGATAGAAACACAG	1606
Db	1989	CAAAAATCTGGTAAAAAATACATATTTAGCTTTTAAAGTCTTAGAAAGAAAGACACAG	2048
Qy	1607	TGTTATTTTTCATCTGGTGCAGCCGTTGCTGCCCTTCCTTTTACATCGAAAACTATT	1666
Db	2049	TGTTATCTATCATCTGGTGCCTGTAGCAGTACCATCTTTTATATGGTAAGTATT	2108
Qy	1667	TGAGACAAAGACGATTTATATATGAATTTGATCGAGTTAATTAATCTATCAATTAATCG	1726
Db	2109	TGGCTGTAAACCGTTTATATATAGAGTTTTCGACAGATAGATAAACCAACTTTTGACAG	2168
Qy	1727	AAACTATGTTATCCCGTAAACAGATATTTTATTTGTTCACTGGGAAGAAATGAAGAAGT	1786
Db	2169	AAAAATTAATGATCCTGTGAACAGTAAATTTATTTGTTCACTGGGAAGAAATGAAGAAGT	2228
Qy	1787	ATATCCTAAATCTATTAATCTGGGAGATTTTATGATTTTGTATGATTCAGTAGAGAACT	1846
Db	2229	TTATCTTAAGCAATTAATTTTAGAGAGAAATTTTAAATGATTTTGTGCACAGTAGAGACA	2288
Qy	1847	CATGAACAACAGTTTAATCGATGTGATTAAGATGATTAATTTGAAAAAAAATGGAAGT	1906
Db	2289	CATGAACAGCAAGTTCAACCGTCTTATTAAGAGATGATTAATTAAGAGGACAGGCT	2348
Qy	1907	ATTAACCGACGAATATTTATTCAAACAGAGATTTCTGATATATTTCCAGATATTTGCAAG	1966
Db	2349	ATTGATCAAGAGTGTTCATTCAAACGGGTATCAAGCTTTGAACCTCGAAATTTGCAG	2408
Qy	1967	TATAAAAAATTTCTCAGTTACAAAGAAATGGAACATATATTAATCAATCAAGAGTGGT	2026
Db	2409	TGCTCAAAATTTCTCATATGATGATATGAACCTTACATGAAGAAAGCTGAGATTTGTT	2468
Qy	2027	ATTGCGACAGAGGCCCGCTACTTTTATGAAATTCATTAATCCAAAGAAAAAAACAATTA	2086
Db	2469	ATCAACACAGGGGCTCCAGCAACGTTTATGATGAGTTTCTAAAGGAAAAAAACATATT	2528
Qy	2087	TTGTTTCTAGACAAAAAAGTAGTGTGACATGTAAATGATCATCAAGTAGATTTGTGA	2146
Db	2529	GTGGTTCCTAGACAGAAACAGTTTGGAGAGCATGTGATATCATCAAGTGTGATTTTGG	2588
Qy	2147	AGAAAGATTTTACAGATATATATTTTATTTATAGAAATATATAGATTTGTTTGAA	2206
Db	2589	AAAGAGTATTTCTGAAATATGAGTTAGATTAATATTTTGAATATCACTGAATTAAGAAAT	2648
Qy	2207	AAAAATATGAGTTTCTAAGCAAC--TAACTTATACATCAAAATTAATTTTTTTGT	2263
Db	2649	ATTATTTAGGAAAAAATATATCTATCTGTAAAGTAATATCAAAAACAATGATTTTGT	2708
Qy	2264	GAAGAATTAAAACAATATGTTGAAAAATTTATGAGATCAAGAAATGAATTAATAAAA	2323


```

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/107,433
  FILING DATE: 30-Jun-1998
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 60/085131
    FILING DATE: May 12, 1998
    APPLICATION NUMBER: 60/051553
    FILING DATE: July 2, 1997
  ATTORNEY/AGENT INFORMATION:
    NAME: Ariniello, Pamela Deneke
    REGISTRATION NUMBER: 40,489
    REFERENCE/DOCKET NUMBER: GTCC-011
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (781) 893-5007
    TELEFAX: (781) 893-8277
  INFORMATION FOR SEQ ID NO: 2303:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1443 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: circular
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
    ORIGINAL SOURCE:
      ORGANISM: Streptococcus pneumoniae
  FEATURE:
    NAME/KEY: misc feature
    LOCATION: (B) LOCATION 1...1443
    SEQUENCE DESCRIPTION: SEQ ID NO: 2303:
US-09-107-433-2303

Query Match      5.9%; Score 415.8; DB 3; Length 1443;
Best Local Similarity 58.3%; Pred. No. 2.3e-61;
Matches 789; Conservative 0; Mismatches 552; Indels 12; Gaps 3;

```

```

Db      582 AGATTTTCAGCATGATTTATTTAAAGTTGTAG-----CAGGGGGGAGATGCTAACTT 635
Qy      554 TTCAACAAAGGAAGTGTGACCAAGTCTTTTAAATCTAACAGTGAAGTTTAAAGCT 613
Db      636 TCGACTACTAGAGGTGTGATGAAGTCTTTTCAATCTTCCAAAGTAAATATACAAATAT 695
Qy      614 AAAGCAATTCGTTTCAGATTTTGAAGTTGTAGATATGATGTAAGCGTTGATTAATTC 673
Db      696 TCGAGAGCTGTCTCTCAGTTTGAACAGATGGAAGTGTATGTAACAGTCAATCTTAATTC 755
Qy      674 ATTCGTTTCTAGCGTTGAAAAAACAATTCACACTGAGTGAACCATGACTTGT 733
Db      756 TTTCGATGTATCTTGACATTAACAGAAATTTGTGATGACGACGACTAAAGCTTGT 815
Qy      734 AACTTTTTCACAAATTTTATTAAGCTTATGATATCATATGATGAAGAAAGCACTTTGGATAT 793
Db      816 GACTTTTCTCAACAATTTTATTAAGCTATGACCATGATGCTGAACGGGTTATGATAT 875
Qy      794 ACTCGAGCGGTAGTCCGGTTAATTAATTTGTGTATAGTTTCTATTTTGTAGTTCCAT 853
Db      876 TATCGGTTCCCTGTAGTGTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 935
Qy      854 TATTCGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 913
Db      936 GATTCAAAAGAGAGGAGGCTCTGCTATTTTGTCTGACAGCGTATTTGGAAAAATGTG 995
Qy      914 CATATTTACATTTCTACAAATTTTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 973
Db      996 CCAATTTCACTTTTACAAATTTTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1055
Qy      974 CTTCGTACGCAAAACCAATGCAAGGAGGAGTGTATTTTAAATGGAAGAAACGATCTCT 1033
Db      1056 ACTCATGAAACAAATATCCATGCA-----GGGTGGAATGTTTAAAGTGTAGTGAATGCCA 1110
Qy      1034 AGAATTAATCTCCAAATTTGCAATTTTCAATGCAAGAAACAAATTTAGAGATTAACCAAGTT 1093
Db      1111 CGATATACAAATTTGTATTTTATACG-GAAGAGACCTTGGAGACCTTACACAGATT 1169
Qy      1094 TTATATGTTTATTTATTTGCGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1153
Db      1170 TTCAATGTTCTTAAGGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1229
Qy      1154 ATTGAAATATATCTCTGCTCAAAAGAGAGATGATGATGATGATGATGATGATGATGATGAT 1213
Db      1230 GTATGAGCACTATATCCCAAGCAAAACGTCCTTAAGTTTAACTGGGCAATACAG 1289
Qy      1214 TCTCTGCGAGTTAGTGTCTGTATATATCAAGACTTCGACGAGTGTGTTGGT 1273
Db      1290 TCTATGCGAGTCAAGCGAGCAAGAGCAAGATCAAGATTTTCGATGAGATGTCTCAATTTAGA 1349
Qy      1274 CTAGCATACATGATTAATTTGCACTATCTGTGATGATTTTAAATTTTAAAGACAGT 1333
Db      1350 TGTGCGCTATATAGACGTTTGGACATTTTGAAGATTTGAATTTTATTTGAAGACAGT 1409
Qy      1334 GAAAGTTGATTTGTTGAGAGAGGAAATGATA 1366
Db      1410 TAAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1442

RESULT 4
US-09-662-254B-23/C
; Sequence 23, Application US/09662254B
; Patent No. 693145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Rawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: US-221C1X1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14

```

; PRIOR APPLICATION NUMBER: 09/086,651
 ; PRIOR FILING DATE: 1998-05-29
 ; PRIOR APPLICATION NUMBER: 60/224,479
 ; PRIOR FILING DATE: 2000-08-10
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: Patent version 3.1
 ; SEQ ID NO 23
 ; LENGTH: 50000
 ; TYPE: DNA
 ; ORGANISM: *Ambacta moorei* entomopoxvirus
 US-09-662-254B-23

Query Match 3.0%; Score 206.6; DB 3; Length 50000;
 Best Local Similarity 43.8%; Pred. No. 7.3e-26;
 Matches 1799; Conservative 0; Mismatches 224; Indels 66; Gaps 18;

QY 2156 TTACAGATATATATATTTTATTTATTTAGAAAATATAGATGATTTGTTGAAAAATTTAT 2215
 DB 19868 TTATATTTACACATTTTGTAGTTAGATTAATATATCTATTAATTTTCGATCAATCTG 19809
 QY 2216 GAAGTTCTAGAGAAC--TAACTTACATCAATAATATTTTGTGGAAGTTA 2272
 DB 19808 TTGTTTGCAGAAACATAGACCAATTAATTAATCTATGACATTTTATTTATTT 19749
 QY 2273 AAACAATAGTTGAAAAATTTAATGAGATCAAGAAATGATTAATAAAGATGCATA 2332
 DB 19748 TGATATATTTTTCACAAAAAATTTATCAATGAAAAAATTAATTTCAAAATGGA 19689
 QY 2333 TTGTATATAGCTTATCATTAATTTTCTCAGATTTTACTGAGAGGATACAGATATAT 2392
 DB 19688 TTACTTAATTTTGATTAATTTTAAATTAATTTTAAATTAATTAATTAATTAATTA 19629
 QY 2393 CATCTCTCAGAGAAATGACACCATTTGTTCTTCAAGAAATCCGTATATTTATTT 2452
 DB 19628 AATTAATTAACAGATTAATTTT---ATTAATTAATTAATTAATTAATTAATTAAT 19573
 QY 2453 AATATATCTCAGATTTATATGTTGAATTTACAAAAGATGAGAAAAATTAAGAAAT 2512
 DB 19572 TGAAGATATGCAATATATATATATGATATATATTTTAAATTAATTTCCAGAAATA 19513
 QY 2513 AGGATATATGACAGATTAATGTTACAGATTAATTTCTATATATATCAGAAAAAATCT 2569
 DB 19512 TTTAAAGTTATATATTTCAATTTAAATTAATTAATTAATTAATTAATTAATTAAT 19453
 QY 2570 -ATTGATATATGCTGTTTGAATTTTATTAAGATATGAGCTTTGAATATCTATTT 2628
 DB 19452 AAATATTAATCATATTTAGATATATCTTATTAACAAAAATAGCAATATTAAGTAAT 19393
 QY 2629 ACAAGATGTTGTTTATGATGAATTAATAAATCATGCTTAAGATTAAGATTTGTTCT 2688
 DB 19392 TACCAATCTATAGATATTTTAAATTTGAATCAATGAATTAATTAAGATCTATTAATTT 19333
 QY 2689 AATGGATTTCCGTTCCACATGATTTTGTGCAATTTCTTTATCAAT---GAAAAAGA 2744
 DB 19332 TTATATATTTAGTAATTTTAAAAAATTAATTAATTTAAAAATTTTGGTAATTTA 19273
 QY 2745 AACAGCTTATTTATTAAGTAATCTAAATGCTCAAGATGAACATTTATTAACAGCAATTA 2804
 DB 19272 ATATATGTTTCTATTAAGTAATGATTAATTAATGAAATCAATCAATTAATTAAGATT 19213
 QY 2805 AGAAAAATATGAAATTTTCAATGATTAATTAATTAATGAAATTTTAAGATATTAAGT 2864
 DB 19212 ATTAATTTTATAGAAAAATTAATTAATTTAAAAAATTAAGTAATTTCTTCAATGTTAAA 19153
 QY 2865 GAAAAAATCAACATCTTCTCTATTTGCTTTACAGATGATCTATGATGATTTGCTAAA 2924
 DB 19152 AAATATATATATCAATTTGATTAATTTTCAAAAAGTATTAATCTATTAATGATTAAT 19093
 QY 2925 TGCAAGAAATTTAGTTTATTTATTTGCTAGAAAGTTAAAAATTAAGAAATTAATCTAAT 2984
 DB 19092 CATATTAAGAAAAATTAATTAATTTT---AAAAAATTTATCAAAATTAATTAATTAAGAT 19035
 QY 2985 TAAAGAAATTTATTAATTAATTAATTAATTAATTTGTAAGATTAATGATTTAA 3044

DB 19034 CGATGATATCTATATGATGTCAAAATTAATCAGATATTTGACATATCAAAATTAATTAATTA 18975
 QY 3045 TTATTTAAATATGACCCCGAAATATTTATTTTAAGACTTGGTGAATTAATTTATTTAT 3104
 DB 18974 ATATTTAGATATACAAAATTTGTAAATTAATTAATTAATTAATTTATGATCAATTAATTT 18915
 QY 3105 CCAGACCAAAAGTATGATTTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTT 3164
 DB 18914 AAAAAAATTTAAATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTA 18857
 QY 3165 TTTTGAATACTAGCTATATTTTAAATAATGAAATTTTATGTTTATTTATTAATGCTATA 3224
 DB 18856 ATATCTATTAAGATGTTTAAATAATCAAAATTTTAAATTTTATTAATTTTATTAATTTTAT 18797
 QY 3225 TTATGTTTGTTCAGTATGACAAATGATTTGTTGAATTAATTTTGAAGATTTATTT 3284
 DB 18796 ACACAATTAATTAAGATTTAAAAAATTAATTAATTTTATTAATTTTATTAATTTTATTA 18737
 QY 3285 GCAGATTTTACTGCTCCATTAATTTGATTAATTTGCAATTAATTAATTTATTAATTTGATTTCA 3344
 DB 18736 GCGGTAAATCAATTTGAATTAATTAATTTTATATCTTTGTAAGTCTACTTAATTTTAT 18677
 QY 3345 TTATTAATTAATTTGATTTAAAAAATTTAAAAAATGATATCTTTTATGTTT--TTAGTTT 3402
 DB 18676 TTATATGTTTATGAAAAATTAATAAATTAATTAATTTATCAATTTGGGTTCAATTAATTT 18617
 QY 3403 TATTAGATATCTGATTTGATTAATTTATTTCAAAATGGAAGATTTGTAATTTTATGACA 3462
 DB 18616 TATTAATTTAAATGAATGATTTTACCAATTTAGATTAAGCAATTAATTAATTAATTAAT 18557
 QY 3463 GACACCTTATAGGACTAGATCTATTAATTAACAGGCGCAAAACAGAGTTGGTGGCTTTA 3522
 DB 18556 GAATTTATTAAGAAAAATTAATTTTAAATTAATTTTATTAATTTTATTAATTTTATTAATTT 18497
 QY 3523 TGAATCTATCTAGCTTAATTAATCACTATTAATTAATTTAGTTCAATTTCCGTTATCTTTGAC 3582
 DB 18496 AATAGATTTATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18437
 QY 3583 TTTATTAATAATTAATTAATTAATTAATTTTCTGTTGCTGCTTTATTAATTAATTAATTA 3642
 DB 18436 TTTAAATTAATTAATTAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAAT 18381
 QY 3643 TTAGTGATGAGAAATTTGATTTATTCGTACAAATTAATTAATTAATTAATTTGTTATGA 3702
 DB 18380 AGGATTTATGAATTAATTAATTAATTAATTAATTAATTTATTAATTTATTAATTTATTAAT 18321
 QY 3703 GATATTAAGTGAATAATTTGCTTGAATTAATAAAGCTAATTAATTAATTTGTAATTTATCTAC 3762
 DB 18320 TTTAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18261
 QY 3763 TTATTTATTAATAATTAATTAATTTTATTAATTAATTAATTTATTAATTTATTAATTTATTA 3822
 DB 18260 CACTAATTAAGAAATTAATTTATTAATTAATTAATTAATTTATTAATTTATTAATTTATTA 18201
 QY 3823 AATCAATTAAGAAATTAATTTATTTATTAATTAATTAATTTATTAATTTATTAATTTATTA 3876
 DB 18200 TATATTAATTAATTTATTAATTTATTAATTAATTAATTTATTAATTTATTAATTTATTA 18141
 QY 3877 TAGAAAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3936
 DB 18140 AAATATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18081
 QY 3937 TCGGAAGTCAATGAGCTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3995
 DB 18080 TATATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 18021
 QY 3996 TTAGGATGTTTCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4055
 DB 18020 AATTAATAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATTTAT 17961
 QY 4056 GCACATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4115

Db 17960 ACTGTTGAGATTTTATAGAAAT-GTATTATATTTTATGTTTGATGATTAATTAATTT 17902
Qy 4116 ATTTATATATTTAGTACTATCTTTCTCAATAGATTTGGAAATATATTAATTTTAA 4175
Db 17901 AAAAAAATATGATTAATTTAATTAATTTGACAAATATATGTTTATATTAATTTTAA 17842
Qy 4176 AAGATATGAGAGCAAAAAATGAAATGATTTAATTTTCAATTATTTGACAAATTTATATGT 4235
Db 17841 TAAATATAAAAATTAATTAATTAATCAAAATAAAAATATATTAATTAATTAATTTACTTAAAT 17782
Qy 4236 CCAAGATATCTTGTAATATGATTTTAACGATATTTTAAACCAACATATATTTTAA 4295
Db 17781 ATACGAAGATATGACTATATATTTAAAAATTTAAAAATTTTCAATTA 17722
Qy 4296 GGTATATCTGTAATGATGGAAGTACTGATCTGAGAAATTTGCTTAACTATAT 4355
Db 17721 ATATTTTATAAAAATTTTATCCAGATATGAAATTTAAAAATTTAAAAATTTAA 17662
Qy 4356 GAAGAACATGGAAGATTAATTAATTAACAAGAAATTAATGCGGCTAGCAGATCTCG 4415
Db 17661 TTACAAAGCAATTCATGACTTTATATCTCAGATATTTTATTTGTTATATATTA 17602
Qy 4416 AAATTTGCACTAGAACATGCAACAGTAAATATATGCTTTGTCGATTTCTGATGCTA 4475
Db 17601 ATATACAAATTTTATATACATTAAGCAAAATTAATGTTTTGAAATTTTAAATATAT 17542
Qy 4476 TATAGAAGTTGCAATGTTGAGAAATGATGATTAATATATACGATATATGCCATAT 4535
Db 17541 ATAAAAATGATTAATATATATTAATTAATATTTGGAATGCTCATTTGCAATTAATAT 17482
Qy 4536 AGCAGATAGATTTTTTGTAGTAGACGAAACGGGTATACAAAGAAAAAGAAATAG 4595
Db 17481 AACATATTAACAATATATGATATTAAGACAAATTAACATATATGTAATTTTATATTA 17422
Qy 4596 TAAATTTTCACTCTTAAACGAGAGAGACTGTAAAAAATTT-----TTG 4641
Db 17421 TATTTTAAATATATCAATTAATTAATTAATTAATTAATTTCTAGAAATTAATTA 17362
Qy 4642 TCGATCTAATATAGAAATTAATGTTGGCAAGCTTATTCAGAGATATTAATTA 4701
Db 17361 AGAATATTAATAAAAAATTAATTAATTAATTAATTAATTAATTTATTTGATATAT 17302
Qy 4702 GATATTAATTTCCAAATTAATTAATTAATTAATTAATTAATTTGCTTTTAAATTTGGAG 4761
Db 17301 TGTGATTAATCTCATGATTTGATATCTGATTAATGATTAATGTCATTAATAATATAT 17242
Qy 4762 GTCTTGAACAATGTAACAAGTATGATGATTAATTAATTAATTAATTTATATTAATGTC 4821
Db 17241 ATATGATGATATATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTA 17182
Qy 4822 ATTGCTAACAGTTCGCTTAATTAATCA-GAAATTCCTATTAATTAATTAATTTAGTCAC 4880
Db 17181 ATTAATAATTTTAATGATTAATTTTGAATTAATTAATTAATTAATTAATTTCTTAAT 17122
Qy 4881 AAGATTTGAGAAATTAACCTTTAAGTTAAAAAGAGATTTAGTCAATTTTGGACAA 4940
Db 17121 TTTATTAATAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA-- 17064
Qy 4941 AGTATTAAGAGAGGTTAAATGTTTAAACAAATGATTAACAAGATTTGTTGAATA 5000
Db 17063 --TTAAAAATATCAAAATATTTGATATTAATAAACAATTAATTAATAAAGAAATGCAAT 17006
Qy 5001 TGAATCTTGGCAATTTAGAGTCTTATCGAAAAAGAAATAGTGAATTCATTTATTA 5060
Db 17005 ATATTAACCTAAATATTAAGTATATGT-TAATATATTAATTAATTAATTTATATACCA 16947
Qy 5061 AGCAAAAGATTTATTAACAAGAAATTTAGTATGTTGATTTGAATTTTGCC 5120
Db 16946 TTTCTATGAAAAATTAATTAATTAATTAATTTATTAATTAATGATCAATTAATTAATTAAT 16887
Qy 5121 TAAATATATGTAATGTTATTAATAAGAAATTTCAAAAGCAGTAGAGTAAAAATGATTA 5180
Db 16886 AGTATTTTATTAATAAATTAACAATATTTGATTAATAAATAAGTAAATATATATAT 16827

Qy 5181 ATTAGCTTATTTGTTCCAGTTTATTAATGATTAATAATTTTAAGTATGTAAGAAAGC 5240
Db 16826 TTTGACAAATATATATATGTAATTAATTAAGAAATGATTAACAATTAATTAATTAATA 16767
Qy 5241 ATTATTAATCAAAATTAATTAATAAATATAGAAATATATGATGATGATGCTGTAGAT 5300
Db 16766 TTAGATTTATTTTATATCTGATTAATTAATAATTTTATATTAATAATTAATTAATAAGTAT 16707
Qy 5301 GATTCCTGTAATAATATGACAGAAATATGACAAAAAGATTAATAAGTAAAAATTTTTTC 5360
Db 16706 TATATTAATAATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 16647
Qy 5361 ACTAATCATAGTGAATATCAAAATCTTGAATCTGGAATTAACGGAATGACACTGAA 5420
Db 16646 AATTAAGTATTAATTAATTTTATGATTAATTAATAAATGAAATTAATCTTAATAATTAAT 16587
Qy 5421 TATATTAAGTTTGTGACCTGATGATGTTGATAGTATGATGATTAAGTAAAAATTAAT 5480
Db 16586 AGTAATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 16527
Qy 5481 TTTAATATTAATTAATAAGTAAAGTATTAATCTGTTGTTGACGTAATTTTCAGAA 5540
Db 16526 ATGCTTAATAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 16467
Qy 5541 AATATTAATTAATTTGAAGTGAATTAATCAAAATATTTGATTTGAAGCAATTAATCCGTG 5600
Db 16466 AATTAATCAATCAATTAATTAATTAATTTGTTAAACAAATCTTAATAATTAATTAATTA 16407
Qy 5601 CAGGACATGCGGAAAAAATTTTATGATTTGATTAATTAATTAATTAATTTTCTCTCT 5660
Db 16406 TTATCTTAATTAACAAAAATCGAATTAATTAATAAGATCTATGAGATTTATATCAATAT 16347
Qy 5661 GTTTGTAACCTATTAATTAAGAAAGATTAATCAATCAATCTTTTCAAGAGATCAATGTTA 5720
Db 16346 GATTAATAAATTTTATTTATTAAT-----TAAATTTCTATAGTAATTAATTAAGACA 16295
Qy 5721 GGAGAAATTTATCTTTTATATCTGATTAATTAAGAAATTAATGATGATTAATTTG 5780
Db 16294 ATTAATAAATTAATCAAGTGAATTAATTTTATGTAATTAATTAATTAATTAATTAATTA 16235
Qy 5781 ACTGAACATCTTATTTTATGAGAGATTAATCAATCAATCAATCTTTTAAAGAA 5840
Db 16234 TGATACCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 16178
Qy 5841 GGTGTTTTTGAATTTGAAATTTGCAAAATTAATTAATTAATTAATTAATTTTGAACAAATA 5900
Db 16177 TAAATATTTTAAGAAATTAATAAATTTATTAAGAAATGATGATTAATTAATTAATTTAT 16118
Qy 5901 TATGTAAGATTTTGAAGT--ATCAATTTGTAAGATTAATCAATGTTGCAAGTATTTT 5958
Db 16117 TAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGTTATTTT 16058
Qy 5959 ATTAATGCTTAATTAATTTTAAATTAAGGAAACAGCTATTTTGAACAATTTTAATTT 6018
Db 16057 TTGACATGATCTCAATTAATTAATTTTAATTAATTAATTAATTAATTAATTAATTAATTA 15998
Qy 6019 TTAAGAAATCTTTAATAAATAATTTATTTTAACTTTGTTAAAGTATTAACAATAATTTCTT 6078
Db 15997 TAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 15938
Qy 6079 TGTCTAAAAATTTTGTATTAAGAAATTTTGAACAAAGTTTAAAAAATAATTAATGCT 6138
Db 15937 AGATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 15878
Qy 6139 TATTAATGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6198
Db 15877 ATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 15818
Qy 6199 ATGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6227
Db 15817 TTAATTAATTAATTAATTTTGTAAATGAAGA 15789

OY	3526	ACTATCCGCTTAATACCACTACATTAATAGTTCAATTCGCCGTTAATCTTGGACCTTA	3585
Db	16856	CAAAATATTGGTTAATTTTAAATAAAATCTAATTTAATATTAATTAATGACATTAATTA	16915
OY	3586	TAAAAATTAANAAGCAACAATTTTTCGTGTCCTGCTTTAATACCGATCTATTAA	3645
Db	16916	TAAATATTATTTTATTTATTTTTCATAGANAGTAATCAATTAATATTATTATTATAT	16975
OY	3646	GTGATCGAGAAATGGTAGTTTATTCGTACGCAATTAATTAATTCCTGTATGAGAT	3705
Db	16976	TAACATCAATTAATTAATTTTATAGTAATATATGATGATTTCTTTATATTATATCTGTT	17035
OY	3706	ATATAGGTGAAAAATTCGTGGATATAAAAAAGCTAATACATATTTGTAATACTACTTA	3765
Db	17036	TTATATCAATTAATTTTGTATTTTAAATAAATCAATATCAATTAATAAACAATATTA	17095
OY	3766	TTATTTTAAATACGAATTCCTTTACATGAATTTTGGCTGTTTAATATCTAGAGAT	3825
Db	17096	ATATATTATTAAATTTTTCATTAATAATTTTAAAGATATATCATATTA--TAAATAAATTT	17153
OY	3826	CAAGTACGAAAGCTAGATTATTAATTTATCAAGAAATATGATAAGTATTGAAAAACA	3885
Db	17154	TATCAAAATATATCTTAATATTTTATTAATTAATCAATTTTGTTTTATTACGTTAA	17213
OY	3886	ATATTTTATTTGATATGGAAT-ATCCGAATTTCACTTAACGGAACTGGCTCGGAAGT	3944
Db	17214	TAAATCATTTATTAATATCAATTCATTAATTAATTAATTTTATATGACATATCATATAT	17273
OY	3945	CATTCAGGCTAATATCATTTTATTAATACGGAATAGTGGGTATTTTACGTATG	4004
Db	17274	CAGATATACAAATCTAGAGTTTATCACAATTAATACAAATAAAGTTTAAATTCATATA	17333
OY	4005	TTTTCTTTTATTAATGTTATTAATAAAAAAGTTATGAGTTAATGGGAAAACAGACTATTT	4064
Db	17334	TAATATTAAATTTTTTTTAAATATCTTAATTTTAAATTTCTTAGAATATA---TATTA	17389
OY	4065	TATTTTACATTAAGCCATATTTTTCATATATGAACAATATGATCCGATTAATATATA	4124
Db	17390	TTCTTTATTTATTTATGATATTTTATTTAGAAATATTTTAAATAATTTTACTATATGTATA	17449
OY	4125	TTAGTACATATCTTTCTTCAATAGATATTTGAAATATATAATTTTAAAAAGATATAG	4184
Db	17450	TT--GCTTTATATACATATATATGTTATTTATGTTATATTAATTCGAAATGTAACATTT	17507
OY	4185	GAGACAAAATAAGATGATTTAAATTTCCGATATTGTACCAATTTAATATGTCACAGA---	4241
Db	17508	CCAAATATTTTTTATATTAATTAATTCATTTTAAATATTAATTTTAAAAATTTGAAAAAGT	17567
OY	4242	TTATCTGATAAATGATATTAACAGTATTAATTAACCAACATAT-ACTAATTTAGCGTTA	4300
Db	17568	TTATTTGCTATATGTTATTAATAATTTTGTATTTTATTTATTAATACAAATATTAATATTC	17627
OY	4301	TTCTCTGTAATGATGGAAGTACATGATTTCTGGAATAATTTGCTTAACTATATACAGA	4360
Db	17628	TCAGATATTAAGTCATTTGAATTTGCGTGAATCTAAATTTTATATTTTATTTAAT	17687
OY	4361	ACGATGAGAAATTAATATTAACAAGAAATTAATGCGGCTACAGATGCTCGAATTT	4420
Db	17688	TTCTATATCTGATTAATAATTTTATTAATAATTTTAATGAAAAATCTTATCTGAATATT	17744
OY	4421	TCGCACTAGAACATGCAACAGGTAAATATATGCTTTTGTGATTCGATGACTATATAG	4480
Db	17748	TTTATTTTAAATAT--ATAGCATATCTTCGTATATTAATTAAGTACAGAAATATTTATTA	17805
OY	4481	AACTTGCAATGTTCCAGAGAAATGCAATGATTAATATPACTGATTAATAGCCGATATAGCAG	4540
Db	17806	TATTTTTTATTTGATTAATTAATTTTATTAATTTTATATCTAAATAATATTTTAAACAT	17865
OY	4541	AAATAGATTTTGTTTAGTACGCAAAACGGGTATACAAAGAAAAAGAAATATGTAATT	4600
Db	17866	ATATTTGCTAATTAATTTAAATTAATCAATTTTTTTTAAATTAATTAATCAATCAAAACATA	17922
OY	4601	TTTCAATCTTAAACGAGAGAGACGTAAAGAAATTTTGTCCAGATCTAATATTAAGAA	4660

Db	17926	AATATATATATACATTTCTATATAAATCTCCACAGATTATAGATTAAGATATCTTTACGCTTT	17985
Qy	4661	ATATATGTTTGTCGCAAGCTTTATTCACGAGATATATAAAGATATATAATTCCAATTTA	4720
Db	17986	ATTTTTCATCTATTTTATCATATATCTGTTTATTTATATTAATTAATTAATTTTTCATATTAAC	18045
Qy	4722	ATTAATAGAGTATTTGGTG--AGGATTTGCTTTTAAATTTGGAGGCTGTGAACATGTAAAC	4778
Db	18046	ATTAATATCATCTATAAATGTTTTGATTTATTAATTAATATCATATTTTTTTTTTATTTTATTT	18105
Qy	4779	ACGTGAGAGTGTGATCTACAGAAATATTAATTAATTAAT--GTCAATCTGAACAGTTTCG	4836
Db	18106	ATTTTAAATTAATATTTGAAATTAACATATTAATTAATTTATTTATAGATATAGATATTAATCTA	18165
Qy	4837	CTTATTTAATCAGAAATTTCTATATAATTAATTTAGTATAGTCAAGATTTGAGCAATTAAC	4896
Db	18166	ATCATTTATATAAATTTATTTAAATATTAATTAATTTTTCATGATATATAAATAATATAT	18225
Qy	4897	CCCTTTAAGTAAAAAGAGAGTTTACTCATTAATTTTGATGCAAAAGTATTAAGAAGAG	4956
Db	18226	CTATTTATTAATAATTTCTATTTCTACATTTAGCTGTTTTTAATATTTATTAATATATAT	18285
Qy	4957	GTTAAATGTTTAAACAAAATGATATCAACAGATTTGTTGGATATGAGTTCTTGCCATA	5016
Db	18286	CTTCAATATTAATTAATGTTTTTATTTTTTTTAAATTTTATTAATTTTTCCAATCTAAAT	18345
Qy	5017	TTAGAGCTTATCGAAAAGAAATACGTAGATATCATTTATTAAGAAGAAAGATATTTA	5076
Db	18346	TATATTTTATATCATGATAGTATCTATATATCTCATGTGTCAGGCAACATTAATTTAA	18405
Qy	5077	TCAGAAGACATTTAGTACGTTGATTTTGA--TGAAATTTTCGCTTAACATATATGTA	5133
Db	18406	ATTAAGTATCATTTGCAATATATATTTTAAATGTTTTTATTTTTTATTTTCTTAGATTTA	18465
Qy	5134	ATCTTATATTAAGAAATTTCAAAAAGCATAG--AGTTAAATAGCATTAATTAATAGTCTTAT	5191
Db	18466	TGTATTTATATCATTAACAAATATATCTATTAATTTCCAAATATTTCTAAATTAATTAAT	18525
Qy	5192	TGTTCCAGTTTAAATGATGATTAATTTTA-----AGTAGTGTATAGAAA	5238
Db	18526	TATTTAAAAAATTACTTTCTTAAATATTTCTACATATTTATTAATTAATTTTGTCTAATAC	18585
Qy	5239	GCAATTTATACAAAATTATMAAAATATAGAAATATTTATGATGATGATGCTCTGTAG	5298
Db	18586	TATTTGGTAAAAATACATTCATTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	18645
Qy	5299	ATGATTTCTGTAATAATATGCAAGGAATATGCAGAAAAGATTAAGAAGTAAAAATTTTTT	5358
Db	18646	CTATATTTTATATTTTCTTAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	18705
Qy	5359	TCACATATCATAGGAGATCAAAATCTGAAGATCATGAAATCAAGAAACGAGTACAGCTG	5418
Db	18706	TAAATTTTATTAATTTCAATGATTTTACAGCATTTAAAAATATTAATATATGATTTACAA	18765
Qy	5419	AATAATATATGTTTGTGACTCTGATGATGTTGTGTAGTAGATTAAGTAAAAATTAAT	5478
Db	18766	TTTTTATTTTTTAACTTAATTAATTTGTGTAAAAAATAATTAATTAATTAATTAATTAAT	18825
Qy	5479	ATTTTAATTTATAAAAAGTAGAAGTATTAATCTGGTG--TTTGTACGCTACTTTTTC	5536
Db	18826	TATTTGATATTTTAAAAATTTCTATGATATTTGTGAATATCAATTAATTAATTTTTTTTGTG	18885
Qy	5537	AGAAAATATAATTAATTTTGAAGGAATATCCAAATATGATTTTGAAGCAATTAATAC	5596
Db	18886	AGTATTTAAATTAATTTTATTTTATTTTATTTTATTTTATTTAGTATTAATTAATTTCTGAT	18945
Qy	5597	C--GTGAGAGACATGGAGAAAAAATTTTATGATTTGTATATATAATTAATTTTCT	5654
Db	18946	TATATTTACATTTTGTATATCTAAATATTTGTAATTAATTAATTTTGTATATGTCAATAT	19005
Qy	5655	ACTCTGTTTGTAAACTATATATAGAAAAAGATACATTAACAGATCTTTTCCAGAGAAATCAA	5714

Db 19006 CTGATATTTGACATCTATGATATCATCGAATTCATATCAATTAATTTGATTAATTTT 19065
QY 5715 TGGTTGAGAGAGATTACTTTTAAATCGCATTTATTAAGAATATAGATAGAGTTAGT 5774
Db 19066 TTAATATATTAATTAATTTCTTTAATGATGATATATACATTAATGATTAATCTTTTG 19125
QY 5775 TATTTGACATGACATCTTTATTTTATAGAGAGATTAAGTATAGTACATTAATCTTTG 5834
Db 19126 GAATTTTATCAATATGATATATTTTATTAACATTTGAAGATATATCTAATTTTATTA 19185
QY 5835 AAGAGAGTGTGTTTTGCAATGAGAAATTTGCAAAACAAGTATAGTATGTTTAAAG 5894
Db 19186 AATTAATTAATTTTCTATTAATTAATTAATTTTATTTGATTTGATTCATATTTTAACT 19245
QY 5895 CAATATATGATGAGATTTTGACATCAATTTGTAAGATATCTATACGTGGCAATGA 5954
Db 19246 CAACATATCTAATAGAGAAACATTAATTAAGTTACCAAAATTTAATTTTATGATATTTA 19305
QY 5955 TTTTATATATGCTTACTATGTTTAAATATAGGAAACAGTCTATTTTGAACAATTTTGA 6014
Db 19306 ATTTTATTAATTTTACTAATTTATTAATTAATTAATTTATGCTATTTATATTCATGATTTCCAC 19365
QY 6015 ATTTTATGAAATCTTTATTAATAAATATATTTAATCTGTTAAAG--TATCTAACAAA 6072
Db 19366 AATTTAAATATCTATGAGATGAGTATTAATTAATTAATTAATTTATTTGTTAT 19425
QY 6073 ATCTTTGCTTAAATTTTGTATTAAGAAATGTTTGAACAAGTTTAAATAATAT 6132
Db 19426 AAGATATATCTAATATGATATATTTTATTTTATTTGTTAATAAATTTAATTAATATAT 19485
QY 6133 TAGGTTATATAGGAAGATATATCATGATCTATTAATAATTTCTATATATGATCTTA 6192
Db 19486 TTAATTTGAATATTAATAAATCTTTTAATTTTCTGAAATATTAATTTAATTTATAT 19545
QY 6193 TATATATATGAGAAATATTTATCTAATGTA-TAGATAGCAATTTGAATTCAGACCTTAC 6251
Db 19546 CATATATATATATATGCAATTTCTTAATTAATTAATTTTAAATATTAATTAATTA 19605
QY 6252 AAAATATATAGATTTCTTCTGATGATGACGGTATGATGATTAATTCGAGAAATTTGT 6311
Db 19606 TAACATATATCTGTTTATTTATTTTATTTTAAATTAATTAATTTTAAATAATTTTATTA 19665
QY 6312 TTATGATATGCGAAGAAAGATAGCGCATCGTTATTTTAA 6352
Db 19666 AAATTAATCAGAATTTAGTAATTCATTTGATTAATTTTA 19706

RESULT 6
US-09-662-25AB-25/c
; Sequence 25, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; FILE REFERENCE: UF-221C1X1
; CURRENT APPLICATION NUMBER: US/09/662, 254B
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086, 651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224, 479
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-25AB-25

Query Match 2.4%; Score 168.2; DB 3; Length 50000;

Best Local Similarity 43.5%; Pred. No. 2.2e-19;
Matches 1662; Conservative 0; Mismatches 2103; Indels 52; Gaps 18;
QY 2466 ATTATATGTTGAATTTTACAAAGATGAGCAAAAATTAAGAATATAGATATATGAC 2525
Db 40165 ATTAATATATATGATCTTCCATTTATTAATTTCTTAATAATATATGATATTTTATTAAT 40106
QY 2526 GAGTTAAATGTTACAGATTTATCTTAATATATCAGAAAAACATATGATATGATCTGT 2585
Db 40105 GTG-TATATATATATATGATCAACATATATTAATCGATGATATTAATAACATGCT 40047
QY 2586 TTAGAAATTTTATAGAAATGATCAGCTTTGAATATATTTTACAAAGATGTTGTTTA 2645
Db 40046 TTCAAATATGATGATATGATGATGCTTTATTAATTAATATATATATATATATTAAT 39987
QY 2646 TTGATATGATATTAATAACATGCTTAAAGATTAAGATTTGCTTATTTGCTTCC 2705
Db 39986 AACTAAATTTTATAGGCAATTAAGAAATATTAATTTTATGAAATTCATATCTAGATGCTG 39927
QY 2706 ACATGATTTTGTGCAATCTTTTATCAATATGAAAAAGAAACAGCTTATTTATTAAGTA 2765
Db 39926 TACAAATGATATGATTAATAAATATTAACATGATTAATAATTTTATTAATAATGATTA 39867
QY 2766 ATCTAAATGTCAGATGACATATTTATACAGACATTTATGAAAAATATGAAATTTTCAAA 2825
Db 39866 TATTAATATTTATACAGAGATTAATATTTATTTATATATATGATTTGATTTATCTTAA 39807
QY 2826 TAGATATCTTAATATGAAATTTAAAGATATATTAAGATGAAAAATCAATCTTCC 2885
Db 39806 TATTAATAT-TACATATGAAAGATGAGAAATATATTTTAAATAATATGATTTGCAAT 39748
QY 2886 TATTTGCTTACAGATGATCTATGATGATGATGCTTAATGCAAGAAATTTAGGTTTTT 2945
Db 39747 TACCAGAGTATTAATTAATATATATATATATATATATTAATTTAATATGAAATA 39688
QY 2946 ATTTGCTAGAAATGTTAAATAATGAAATTAATCTAATTTTAAAGAAATTTTACTAATAA 3005
Db 39687 TTGTATATGATCTATTTAGAAAAATTTCAATATATATATTAATATGACAGATTTGATAT 39628
QY 3006 ATTAATAATGTTGATTTTGAGATGATATGATGTTTAATTAATTAATATATGACCCGGA 3065
Db 39627 ATAGTTTA--AGATCAGATCAAGAAATTTAATTAATTAATAAATAATATATATATATAT 39571
QY 3066 TATTTATTTTATAGATCTTCTGATGATATTTTATTCAGAGCAAAAGTATGATTT 3125
Db 39570 TTAATTAATAATGCAATTAATATATATGATCAATTTTCCAGAAATTCATGATTAATTA 39511
QY 3126 TTAATTAATTTTATGATTAATTTTATTTTATTTTATTTTATTTTGAATACTAAGCTAATA 3185
Db 39510 TAATTAATAAATTAAGATTTTATGATTAATTAATTAATTAATTAATAAATTAATATATAT 39451
QY 3186 TTAATAAATGAAATTTTATGTTTTTATTAATGATCTAATATATGTTTTGTTCAATATGTC 3245
Db 39450 ATGATTAATATGAAATTAATATATATTAATAAATCATTAATTTTAAATAATTAATAAGAT 39391
QY 3246 ACAAGATGTTGTTGAATTAATTTTGAAGATTAATTTGCAATTTTACTGCTCCATA 3305
Db 39390 ATATTAATGAAATTAATAAATTAATAATTAATTAATAAATGATCAATGATGATTT 39331
QY 3306 ATTTGATTTATGCAATATGATTAATTTTATTTTATTTTATTTTAAATATTTGATTAATAA 3365
Db 39330 TTTTATATTTTATTAATTAAT-TATTAATAAATAATGATTTTATTAATTTTAAATAAT 39272
QY 3366 AAATTAATAAATATGATCTTTT-TAGTTTTTATGTTTTTATAGATATATCTCATGATGTA 3424
Db 39271 AATGAAGATTTTATTTTATTTTAAAGAAATATATTAATTAATATATATTAATTAATATA 39212
QY 3425 TATTAATCAAAATGGAAGATATGATTTTATGACAGACCTTATAGACATGATGATGTA 3484
Db 39211 TTAATCATGATGATTAATAAATTAATTAATGATGATTAATTAATTAATAAATAATATA 39152
QY 3485 TCTTATTAACAGCGCTCAAAACAAGTTGTTGCTTATGACATATCTTAAATATAC 3544

Db 39151 GATATTATATGATGATTATAGAAAAAATAGCATGTATATAAAAAATTAAAAACAAAA 39092
Qy 3545 CACTCAATATTAGTTTCAATTCCGTTAATCTTGGCACTTAATAAAAATTAATGACACA 3604
Db 39091 ATATCAGACGATGAATACCTTATCTAATTTATTTATCAAAATGAAAAATTTTAAAAAAA 39032
Qy 3605 ATTTTCTTCTGTCCTGCTTTTATACGATCTATTTAAGTGATCGAGATTGGTAG 3664
Db 39031 ATTTAAAGATTTATTAGCAGTGAAAAATATAGATATGATTCCTATATTAATGTTTAAAT 38972
Qy 3665 TTTATCGTAGCAATTTAATTATATGCTTTGTTATGAGATATATAGTGGAATTTGCG 3724
Db 38971 ATTTAAGAAAAATTTAAAAATATTTTAAAAAGAAATTTAATAGAAATTTTAAAAAGA 38912
Qy 3725 TTGGAATAAAAAGCTAATAGTATTTTGTATCTACTATATTTTAAATCTAGATTT 3784
Db 38911 AATGATATATGATTTAGTTAATTAATTTAATGATTTATTAACAAATATTAATAATCTATT 38852
Qy 3785 GCTTTACCAATGAATTTGGCTGTTTATATTTCTAGAGATCAAGTAAAGAACTAGATTT 3844
Db 38851 GATGAAAAAATATATGATATTAATTTAACAAGATGAATTTAATACATCATCATCT 38792
Qy 3845 TATTTATTTCAAGAAATATGATTAAGTATTAAGAAACAAATTTTATTTGGATATGG 3904
Db 38791 TTGATGTTTTAATATCTCCATATCAAAATCTTAGAA---CTTATATGATGATATAGAT 38735
Qy 3905 AATATCCGAATATTCAGTTAOGGAACTTGGCTCGAAGTCATTCGGCTATATCATTT 3964
Db 38734 ACTATGATGAACTATATATGATTAATATGATTAACAAATATTTAA--ATCTATATCAGG 38677
Qy 3965 TTTTATTAATCAAGATATGTTGGGTTGATTTTATCTGATGTTTTCTTTTATTTATGTTAT 4024
Db 38676 AAATGAAAAACATACAGATCATTTATTAACAAATGTAATTTCTAGATTTTTCGCAATTAT 38617
Qy 4025 AAAAAAAAGTTATGAGTAAATGGGAAACAGCATATTTTATTTTATTTATCATCATTAAGCAT 4084
Db 38616 CTGATGTTAGTAAATTCGATTTGATGAAGAAAAATATGATTAATTTTATTAATAAAATCAAGATGCTA 38557
Qy 4085 ATTTTTCATATATGAACATATAGATCCGATTTATTTATTTATTTAGTACTATCTTTCTTCTTC 4144
Db 38556 TTTTAAATTAATAATTAAGATTAAGAAAGTATTTTATTAACCTACGTCGAAGATATATCTA 38497
Qy 4145 AATAGGATTTTGGAAATATATTAATTTTAAAAAGATATGAGACAAAAATGATGATTT 4204
Db 38496 AAAATATTAATTAATTAATGATGATTTTCAATTAATTAATTTAACAATATATGAAATTT 38437
Qy 4205 TAAATTCAGTTATTTGTAACAATTTATATATGTCGAAGTATTTCTTGAT-----AAATGTA 4258
Db 38436 TAAATAGTGTGATGTAAACAAATTTGATGATTAATAAAATATTTGATTTTATCAAGAAATTA 38377
Qy 4259 TTTAAGATTTATTAACCAACATATCTAATTT-----AGAGTTATTTCTGTAATAG 4312
Db 38376 TTTAAATTTTATTAATAATTAATTTTACAGAAATTAATTAATAATTTTAAACGGAAT 38317
Qy 4313 ATGGAATGATCTGATGATTTCTGAGAAAAATTTGCTTAAACTATATGAGAAACGATGAGAA 4372
Db 38316 ATGTAATCTGATTAATTTTAAAAATATTTATTAATGAAAGTATTAATTAATTAATAATA 38257
Qy 4373 TTTAATATTTACAGAAAAATTAATGCGGCTAGACAGATGCTCGAAATTTTGGACTAGAAC 4432
Db 38256 CAGATTATATTTAATTAATTTAATAATCTAATAATTTCTTCACTGAAAAATATGATTAATTTCAA 38197
Qy 4433 ATGCAACAGTAAATATATTTGCTTTTGTGCAATTTCTGATGACTATATAGAAATTCGCAATGT 4492
Db 38196 AGGTATGATCTAGATTAAAAATTAATTTATGATTAATTTTATGATTAATAACAAATAGTA 38137
Qy 4493 TCGAGAGATGATGATTAATTAATCTAGTAAATGCGATATGCGAGATGATGATTTT 4552
Db 38136 TAACTTTATTTAATACATATTTAGCACTAATTAATGCTAAATATCTGAATTTGCGCGTG 38077
Qy 4553 GTTTATGATGAGAAAAACGGGTATACAAAGAAAAAAGAAATGTAATTTTCATGTCTTTAA 4612
Db 38076 ATATCGAAAAAATTAATAATATTTAGCGAAAAATATTAATACTATCTATCGAGATTT 38017

Qy 4613 CGAGAGAAAGACCTGTAAAGAAATTTTGTGACGATCTAATATAGAAAAATAATGTTGGT 4672
Db 38016 TTTAATGACGAAAAAGCAAAAAATAGATTAATAATTAATAATTTTAAATAATCTGTTT 37957
Qy 4673 GCAACCTTATTTCAAGAGATTTATTAAGAAATTAATAATTTCCAAATTAATAATAGAGTA 4732
Db 37956 TGTATCAATCTAATAATCCAAATATTTAAGAAATATTAATAAGTTATATCTTCAAGAGCAA 37897
Qy 4733 TTGATGAGAAATTTGCTTTTATTTTAAATTTGAGGCTTGAACAAATGTAAACAGGTAGATTGG 4792
Db 37896 GTTTGCCAAATTAATAATTAATAATTAATTTATTTATGATTAATAATTTATCTGATTAATAAA 37837
Qy 4793 ATATGAGAAATATTTATTAATTTATGTCATTTCTGAACATTCGCTTATTAATCGAAAT 4852
Db 37836 TAAATATGAGAAATTAATCAAAATTTTGAATTTGATGAGATTAATAATTAATAATCAATCG 37777
Qy 4853 TCTCTAATAATAATTTGATTTAGTCACAAGTTGAGAAATTAACCCCTTAAGTTAAAAA 4912
Db 37776 ATTTATTAATAATTAATCAATTTAG--ATGAATTTAAGAAATTTATATCTAATAACAAAAAC 37719
Qy 4913 GAGAGTTTATGATTTTGTGATGCAAAAATTTATTAAGAGAGTTAATGTTTAAACA 4972
Db 37718 AGAAATATGTCAGAAATTTGATGAAAAAATAATGCAATTAATGATTTATGTTACTTT 37659
Qy 4973 AAATGATTTCAACAGATGTTTGAATGATGATTTCTGCCAATTTTAGAGCTTATCGAA 5032
Db 37658 ATTAGAATCAAAAAATGAAAAATGATGAGAAATATATGATTAATTAATATATCTTAA 37599
Qy 5033 AAGAAATACGTATGATTCATTTATTTAAGCAAAAATTTATCAAGAAAGCAATTTAG 5092
Db 37598 TAAATTAATTAAGAAATTTAACAATAATGATTAATTAATTTTATTTAATTTTGAAGAAAA 37539
Qy 5093 TTAAGTTATTTGATGAAATTTTTCCTTAACTAATGTAATGTTATTAAGAAATTTTC 5152
Db 37538 AATGAAAAATGCTTGATTAATTAATTAATTTATTTAATTTATTTATGATTAATGATTAATA 37479
Qy 5153 AAAAGCAGTAGAGTAAAAATGATTAATAATGTTATTTGTTTCCAGTTTATTAATGTAA 5212
Db 37478 AATTAATCTAATAAGAAATTAAGATTTTATTAATAATAATTTATGTCAGAAATTTATCC 37419
Qy 5213 TAAATTTTATGATGTTATGTAAGAAAGCATTTATTAACAAATTTATTAATAATTAAGAAAT 5272
Db 37418 AAAATATTAATA--ACTGAAACGAAAAATCTTAAAGGTTATGCAAAACATTAATAATA 37362
Qy 5273 ATATGATGATGATGATGCTGCTGATGATGATTTCTGTAATAATATGCAAGAAATATGACAGA 5332
Db 37361 GATTAATTAATTAATAATTTCTTTGGAATAATGCTCTAGAAATTAATTAATTAATTAATTA 37302
Qy 5333 AAAAGATTAAGATTAATAATTTTTCATCTAATCATAGTGAGATTCAAATGCTTAGAA 5392
Db 37301 GAAATCTAAGTTCTAATAAAAGTAAATCTAATATGTAATGTTGCTAATTTAATAAA 37242
Qy 5393 TCAATGAAATTAAGCGAGATACGCTGAATTAATTAATTTAGTTTGAATCTGATGATGTGG 5452
Db 37241 --ATATATTAATAATTTGTAAGAAAGAAAGAAATCCATTAAGACACCTGATTTGTTCAA 37184
Qy 5453 TGAATGATGATGTAAGTAAATAATTTATTTATTTAATAATAAGTGAAGTATTAATTC 5512
Db 37183 GAAATTAATAAAACCAATTTAATTAATGATTAATAATGATTTTAAATAATAATTAATTAAT 37124
Qy 5513 TGGTGTGTTGTAAGTACTTTTCAAGAAATATTAATAATTTTGAAGTAAATTAATTAATTA 5572
Db 37123 AAATCAATTAATCAATTAATAATTAATAATTAATAATAATAATAATAATAATAATAATA 37067
Qy 5573 TATTTATTTTGAAGCAATTAATATACGTCGACGAGACATGCGAGAAAAAATTTTATGAAATTT 5632
Db 37066 TTTGAATATGACATTAATTAATGAAATGATTAATTAATTAATAATAATAATAATAATAATA 37010
Qy 5633 GTATTAATAATAATTTTTCATCTGCTGTTGTAATCTATTAAGAAAGATACATTAAC 5692
Db 37009 -TATTAACGGAAGATATTTCTTATTTATATATCAATTAATAATAATAATAATAATAATA 36951

OY		2817	ATTTCCTCAAATGAGTTA-TCTAAATATGGAAAATTGAAGATATATAAAGCGGAAAAATCA	2875
Dd		18032	AATTATAAATTAAATTAATTTATAAATTATAATTCACGTATTAATATTAATATCTAGATA	17973
OY		2876	CATCCTCCCATGTGCCTTACAGATGACTCTATGAGTAGAATGCTAAATGCAAGAATT	2935
Dd		17972	AAGAAATTTATATATGTTTTATTTCAATCATGATATATACCTATTTTAAABAAAT	17913
OY		2936	TAGCTTTTATTTTGCTAGAAAAGTTAAAAATAGAAAAATTAATCTAAATTTTAAAGAAATTA	2995
Dd		17912	TATGTAAATAACTGCAAAAAAATTAATATTAATATATCTAATTTATTTTAAAAATTTTG	17853
OY		2996	TT-----ACTAAAAATTAANAATAGTGAATTTTGAGAGTAATGATATGTTTAAATTA	3047
Dd		17852	ATTGCTCTAAATTTAAATATATATATAACAATAATTAATTTGTATCGAAGATGATATTAAT	17933
OY		3048	TTTTAAATATGACCGGAATTTTATTTTAAAGTACCTCGTGATTTATTTTATTTCCA	3107
Dd		17792	TTTATFACCTTBATATATATATTTTTTTTTTANGTAATTAATTAATTAATTTTGATTTTAAT	17733
OY		3108	GAGCAAAAGTATGATTTTATTAATTTTATATGATTTAATTTTATTTGCATATPAAAATTT	3167
Dd		17732	TTAATACATPAATAAATAAACCTAACCTACATCATATATATATATCTAATATATAAATA	17673
OY		3168	TTGAAAACTAAGCTAATATTTAAAAATGAAATTTATGTGTTTTATATATAGCTATATTA	3227
Dd		17672	TAAATATATATACMAAAATAGMAAAATTTATTCAAATATCACAAATTAATCCATATATA	17613
OY		3228	TGTTTTGTTCAGATGACACAGATATGTTTGTAATTAATTTTGAAAGATATATTTGCA	3287
Dd		17612	CAGATTAACCTTAAATATATATAA- TATATAGACCATTGCCATATATATATATTTATTTTA	17554
OY		3288	GATTTTACTGCTCCCATAATTTGGATATTTGCATATAATGATATATATTTGATTCATT-	3346
Dd		17553	AACTTTTAAAAATATATATA TTGAAAAAATATATCAATTTATATATATATATATTCAAAT	17494
OY		3347	-----TATTAATATGATTTTAAAAAATTA AAAAATATATGATCTTTTATGTTT	3394
Dd		17493	GGTATATATATAAGTAAAAATATCGTTCAATTAATACCTAAAAACAAGATCTGTATTTGTT	17434
OY		3395	TTTATGTTTATAGGTATATCTGATGTATATATTCAAAATGGGAAAGATATGTGATT	3454
Dd		17433	TTTGGTGTTATATCCAGAAATATTTGTCATATCAATATTTACATTAATATATAAATATA	17378
OY		3455	TTTAGACAGACACTTATAGACATAGACTATCTTATTAACAGGCGTCMAAAAGTTTGGT	3514
Dd		17373	CTAATATATATTAATCGGTAAAGAAACATATATATAATTAACAATCTTTATTTATTTTA	17314
OY		3515	TGGCTTATATGAACTATCTACGTAAATATPACCACTAATATATAGTTTCAATCCGTTAAT	3574
Dd		17313	ATATATATTAATTTGATTAATATATGACATATAGATGTAAATTTTATATACAGAAAAATACATA	17254
OY		3575	CTTTGCACTTATAAAAAATTA AAAGCAAAATTTTTTCTGTGTGCTGCTTTATPAC	3634
Dd		17253	TACGCAAAAGAAACATATTAATATATGTTAAATTTTTCATATCTATANTAGATATG	17194
OY		3635	GATCTATTTAATAGTATGAGATGGAATTTGGTATTTGCTGACATATATTAATTATGCTT	3694
Dd		17193	ATATTTAATATATATGTTATTTAAATATTTGGAACATTAATAACTAATATTTCTAAATCATAT	17133
OY		3695	GTTATGAGATATATAGGTGAAAAATTTGCTTGATTA AAAAGCTAAATVAGTAATTTTGT	3754
Dd		17133	ACATTAATTTTATTTCTATTTATTTAAAGTAACAAATATATAATATATGTTAATATATA	17074
OY		3755	AATACTACTTATATTTTAAATCTCGAAT - GCTTACCATGAAATTTTGCGTTTATAT	3813
Dd		17073	AATACACACGATACATATTTATATTTATTAATAAATAATATATATCATATTTTATTTTCA	17014
OY		3814	ATTCAGAGAAATCAAGTAAGAGCTGATTTATTTATTTATCAAGAAAGTATGATTAAG	3873
Dd		17013	TTTATATATATTTCTATATACTTAATCTTAATTTATTTATTTTATATATATTTATATACG	16955
OY		3874	TATTAAGAAACAAATATTT---TATTTGATATAGAAATATCCGATATATCAGTTACGGG	3929

Db	16953	TATTCAGTATCAATTTTATTCATTGTTATAGAAATTGAGTAATTTCCATATATAT	16894
Oy	3930	ACTTGGCTCGAAGTCATTCAGGCTATATATCATTTTATTAATTCAGAAATAGTGGG	3989
Db	16893	ATTCCATTTTCTATTAATTAATATTTATATATTTTATTAATAACATCGTATGAT	16834
Oy	3990	TTGATTTTACTGATGTTTTCTTT-----TTTATATGTATAAAAAAAAGTTAG	4038
Db	16833	TTTTCTTTAAATGTTTTATCATTTAAAGTTGTTAAATATATTTTAAACGTGCATATTT	16774
Oy	4039	GAGTTAATGGGAAACAGCATTTTATTTTACATTCATTAACATTTTTCATATATG	4098
Db	16773	GTTTTATTTGTTAATTTAATTTTATTCATTAATAAGTTCAATTTGTAATTTCTAAA	16714
Oy	4099	AAACAATAGATCCGATTAATTAAT-TATATTAAGTACTATCTTTTCTTCATAGGTATTTGG	4157
Db	16713	AAATATGTTTTATTTATTAATACGTTTATCATTTAATAATATTCGTTAATCCACAACAT	16654
Oy	4158	AATATATTAATTTTAAAAAGATATGAGACAAAAATGAATGATTTATTTCAGTTAT	4217
Db	16653	AATGATATGTTGTTTAAATACCAAAAGTTTGATTAATCATATTAATTTAATTAATTAAT	16594
Oy	4218	TGTAACCAATTTAATATGTCAGAAATATCTGTAAATGATTAATTAACAGTATTAACCA	4277
Db	16593	ATTTTCAATGTTAAATATTTTAATATAGTATTTAATTAATTAATTTTTCATTTAT	16534
Oy	4278	AACATATACAAATTTTGAAGGTTATCTCGTAATGATGAGAGTACTGATATTCAGAA	4337
Db	16533	AGTATATTAATTTTTTTTGAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT	16474
Oy	4338	AATTTGCTTAATCTATATGAGAGACATGAGACATTAATTAATTAACAGAAATTAATCG	4397
Db	16473	TAGATATTAATGATATTTTAAATATAGAAAAATTTGTCAATATGTAAGAAAAATATAG	16414
Oy	4398	CGGCTAGCAGATGCTCGAAATTTGCGACATAGAACATGCAAGTAATATATTTGCTTT	4457
Db	16413	AATTAAGTAAAAATGATATGATGTTTAACTTGCATCTTTACATATATGTACAAAAATGTT	16354
Oy	4458	TGTGATTCGATGACATATATTAAGTGTGA-----ATGTTGAGAGATGCATATATA	4511
Db	16353	CGTATTAACAAATTAATTCATATCATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	16294
Oy	4512	TATTAAGTATATATGATCCGATATAGACAGATAGATTTTGTGTTAATGTAAGCAAAAAACGG	4571
Db	16293	TTGCTCATATGCTTAATATGATATTTATGATCAAGTTTTTAAATCAATATTTATGACTTATA	16234
Oy	4572	GTATACAAAGAAAAAAGAAATGTATTTTCAATGCTTAACAGAGAAAGACCTGTATA	4631
Db	16233	GTAAAGATGATTAATTTGAATATATCAATATATATTAATTAATTAATTAATTAATTAATTAATTAAT	16174
Oy	4632	AGATTTTTCGCGAGATCTAATATAGAAAAATATGTTGGTGCACCTTAATTCAGGAA	4691
Db	16173	ATTAATTAATTAATTAATTTAAACAAAAAATATCTTAATGAAAAATTTAATTAATTTTA	16114
Oy	4692	TATATATAAAGATATAAATTCGAAAT-----AATATAGAGATTTGGTGAAGTTTG	4746
Db	16113	T-TTATTAATAATAGTATTAACGAAAAATTTTCAAAATATATATATGTTTTTAAATTAAT	16055
Oy	4747	CTTTTAATTTGAGGCTCTTGAAACATGTAAACCGTGTAGTACTGATCTAGAGATAT	4806
Db	16054	AATAGTATTAATACATGATTAAAAAATGTTTTATTTGATACAAATTTTATTAATAATTAAT	15995
Oy	4807	TATTAATTAATATGCAAT-----TCGTAAGAGTTCCGTTTATTAACAGAAATCTCTG	4857
Db	15994	TATATATTTTCAATAAATAATATATTTCCGTTTATTAATTTTAAATTAATTAATTAATTTTAAAT	15933
Oy	4858	ATTAATATAATATGATTAAGTACACAGATTTGAGATTAATCCCTTAATGTTAAAAAGAG	4917
Db	15934	ATTGATGATATATGTTTATGATATACATTTGAAAAATTAATAATTTTAAATATGAAATTTCCAG	15875
Oy	4918	TTTAG-TCATTAATTTGATGCAAAAGTTATTTAAAGAGAGTTAAATGTTTAAACAAAT	4976

Db 15874 TCTAGAGTAAATTTTACAAATACACATGACCATGAAACAAAATTAATTTCTAAATTA 15815
Qy 4977 GATTCAGACAGATTTGTTGGATATAGAGTCTTGCCAAATTTAGCTCTATGCGAAAGA 5036
Db 15814 ATAAAAAATGATATGATATGATATTTTATTAATTTG--ATAAAAATTAATTTTGGCGAAT 15758
Qy 5037 AATACGATGATATTCATTTTATTAAGCGAAAGATTTTATCAAGAAAGCATTTAGTTAC 5096
Db 15757 ATATATGATTTATATATATATTCATATGATATTAATAAAAAAAGAAATTAATAGTAAC 15698
Qy 5097 GTTGTATTTGATGATAATTTGCGCTAACTATATGTAATGTAATATTAAGAAATTTCAAA 5156
Db 15697 GCTGAATATTAATTAATTTATATTTGTAATGTTAAGCACCAAAACAACTAATATTA 15638
Qy 5157 GCAGTGAAGGTAAATATGATAAATTAAGTATGTTCCAGTTTATATATGATATTA 5216
Db 15637 ACTAGATGAAAAATTCGATGATATTAATGATGATTAATTAATTAATTAATTAAT 15578
Qy 5217 TATTTAAGTATGTTGA-----TAGAAAGCATTTATTAATCAAAATTAATAA 5261
Db 15577 GTTATTAACATATATATATATATTAAGAAATCTATTAATAATTAATAATTAATAA 15518
Qy 5262 AATATAGAAATATTTATGATATGATGCTCTGATAGATTTCTGCTAAATATGCAAG 5321
Db 15517 TATATTTAACTATGATTTTAAAGATGCTTTGAGCGCATTAATATCTCATATATGT 15458
Qy 5322 GATATGCGAAAAAGTAAAAAGTAAAAATTTTTCCTATATCAATGATGAGATCA 5381
Db 15457 TATTTGTAATTAATTAACATATTTATGTAATTAATTAATTAATTAATTAAT 15398
Qy 5382 AATGCTAGAAATCATGGAATTAAGCGAGACGCTGAATATATTTGTTGTTGACTC 5441
Db 15397 ATATATATGATATATATTTATTAATAGTAAACCCAAATGATTTTAATGATATCA 15338
Qy 5442 GATGATGTTGTTGATGATATGATATGATAGAAAAATTAATTTATATTAATTAAGTGA 5501
Db 15337 CCATATATATTTTAAATGCTATGATATGATATTAATCAAAAACTATTAATAAATAA 15278
Qy 5502 AGTATTTATCTGTTGTTGTAACGCTATTTTCAAGAAATTAATAATTTTGAAGTG 5561
Db 15277 GATTTTAAATTAATGATGATGATTTTATTTTGAATATGATTTTAAATTAATTT 15218
Qy 5562 AATATCCAAATATGATTTTGAAGCAATTAATCCGTGAGACATGGGAGAAAAA-- 5619
Db 15217 TATTTATATCGAATTTTATTAATTAATTAATTAATGACATATATGCCCCATTAATA 15158
Qy 5620 -----ATTTATGAATTTGTAATTAATTAATTAATTTTCTACTCTCTGTTG 5665
Db 15157 AATTAATATAGATTAATGAAAAAACAATGCAAAATATATATTTTAAATATATCTGTTAT 15098
Qy 5666 TAAACTATATAGAAAAAGATACATAACAGATCTTTTCAAGAGAAATCATGTTAGAGA 5725
Db 15097 TTTCTAGTTTATTTATCCGATTAATAATTAATTAATGCGCTCATGAACTTATATGATT 15038
Qy 5726 AGATTTACTTTTAACTGATATTTAAAGAAATATAGATAGAGTTAGTTATGCTGA 5785
Db 15037 AATCTTAAGTTTATTTTAAATTAATTAATTAATTAATGCAATATATATTAATTAAGT 14978
Qy 5786 ACATCTTATTTTATAGAGAGATATCTAATGATAGTAAATCTTT-----T 5834
Db 14977 AATTAATTAATAGATTAATGATATTTCTTATTTTGAATAATATATATATTAAGT 14918
Qy 5835 AAGAGAGGTGTTTGTGCAATGGAATTTGCAAAAACAAGTATGATTTGTTAAG 5894
Db 14917 TAAAAAGAGAGTGTGACCAATTAATTAATAATTAATAATTAATTAATTAATAA 14858
Qy 5895 CAATATATAGTATGAGATTTTGAAGTATCAATTTGTTAAGATCTATGCTGGCAAGT 5954
Db 14857 ATTTGAAAAATATATTTGTTAAGGATATTTAAAAATTTTGAATTAATAATTC 14798
Qy 5955 TTTTATTAATAGCTTAAATGTTTAAATAGAAAAACAGTCTATTTTGACAAATTTTGA 6014
Db 14797 ATTTTATGATTAATCTGATATCACTAATAATTAATAATAATAATAATAATAAT 14738

Qy 6015 ATTTTGAATCTTTTATTAATAATAATTAATTTTAACTGTTAAAGTATCAAAAAAT 6074
Db 14737 ATATATATCATATGTTGATATATGTTGATTAATATGACATCAATTAATAATTAATA 14678
Qy 6075 TCTTGTCTAAAAATTTTGTATTAAGAAATTTGTTCAACAAATTTTAAAAAATTTA 6134
Db 14677 AAAAAATTAATAATATATGTCAGAGATATATTAATTAATTTTAAACAAATCAAAAA 14618
Qy 6135 TGGTTATTAATAGAAATATCATGATATCATATAGTAAATTTCTAATATGTAACATA 6194
Db 14617 AGACGAAAGAAATTAACAATTAATATCGAATCTGAAGATTTAAGATGCAATTAATGA 14558
Qy 6195 TATATGTAAGAAAAATTTTATCTAATATGATATGATGATGATGATAA 6241
Db 14557 AATGATGTAATTAATTAATGATATATGATATCAACAAATTAATTAATGAAA 14511

RESULT 8
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marke
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 2.3%; Score 158.4; DB 3; Length 1830121;
Best Local Similarity 56.4%; Pred. No. 1.5e-17;
Matches 359; Conservative 0; Mismatches 266; Indels 11; Gaps 3;

Qy 747 AATTTTAAAGCTGATCATATCATGATGAACGACTTTGGATATATCTCGAGCGGTA 806
Db 924401 AATCTGCTAAACTATCTCTGATATTTGAAACGAACATGATATTTGTTGCTCC 924460
Qy 807 GTCCGGTATTAATTTTGTGATATGTTCTATTTTGTAGTTCATTTTCGTAGAGAT 866

```

Db      924461 TTAGCTATTATTATATTTTCCCAAGTGTGCTTTATCTGTATTTTGGACGATTAAGAT 924520
QY      867 GGTGACCGGCGATTTTGTCTCAGAAACAGATTTGACAGATGAGACGATATTTACATTC 926
Db      924521 GGTGATATGCTATTATTTATGAGCATCTAGAAATTTGGTCGTAATGAGAAAAACATTTACTGT 924580
QY      927 TACAAGTTTCAGTCAGATGATGATGCTGAGAGACGCAAAAAGACTTGTCTAG---- 982
Db      924581 TTGAATTTAGAAACATGCTGTAATTTCTAAAGAAAGTCTAGATGAAATGCTTAAGACA 924640
QY      983 ---CCAAAACAGATGCAAGGGGTGATGCTTTTAAATGGGAAAAACATCTGAAT 1039
Db      924641 GATCCTGAAGCAAGCTGATGAGAAAAAGATTTTAAATTAAGATGATCTGAAT 924700
QY      1040 ACTCCAATTGACATTTGATACGCAAAAACAATTAGACGATTCACAGATTTATTA 1099
Db      924701 ACTAAATGCGTCTTTATTTGCGC-AAAACAGTTTGAATGATTAATCTCAATTTTAA 924759
QY      1100 TGTATTAAATGCGATATGATGCTAGTGTGACAGCTGCACTACAGTTGATGAAATTTGA 1159
Db      924760 TGTATTAAAGGTAATGATGATTTAGTTGGGCTCGCCCAATGTAATGATGAGCTAGA 924819
QY      1160 AAAATATACCTCTGCGCAAAAAGACGATTTAGTTTAAACGAGGATTAAGGCTCTG 1219
Db      924820 GCGTTATGAAGAAACGTAGATTATTACTTAATGCAAGACCGGATGACAGGCTTTG 924879
QY      1220 GCAGGTAGTGTGCTAGTAAATATCACAGACTTCGACGATTCGCTTGAATTTAGC 1279
Db      924880 GCAAGTCAGCGACGTAATTAATTC--GATTAATATCTCGGTTTATTTGATTTCTTG 924936
QY      1280 ATACATTTGATTAATTTGATGATCTGTCAGATATTTAAATTTTAAAGACAGTGAAGT 1339
Db      924937 GATATGTAATAAATTTGCTACTTGTGATGATGATTTGATTTTAAACAATGATGT 924996
QY      1340 TGTATTGTTGAGAGAGGAAAGTAAGTAAAGTATAT 1375
Db      924997 GGTTTTAAATCGTATGAGCTTATTAACGTTATTT 925032

```

```

RESULT 9
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
;           Mark D. Adams
;           Owen White
;           Hamilton O. Smith
;           J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
;           the Haemophilus influenzae Rd Genome, Fragments
;           thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787

```

```

; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match      2.3%; Score 158.4; DB 3; Length 1830121;
Best Local Similarity 56.4%; Pred. No. 1.5e-17;
Matches 359; Conservative 0; Mismatches 266; Indels 11; Gaps 3;

QY      747 AATTTTAAAGCTTAGTCATATCATGATGAAACGACTTTTGATATACCTGAGACGGTA 806
Db      924401 AATCGGCTAAACATATCTTCTGTAATTTGAAACGACCATGATATGTTGTTGCTCC 924460
QY      807 GTGCGTTAATTTATTTTGATATAGTTTCTATTTTGTAGTTCCAAATTAATCGTAGAT 866
Db      924461 TTAGCTATTATTATATTTTCCCAAGTGTGCTTTATCTGTATTTTGCAGTAAAGAT 924520
QY      867 GGTGACCGGCGATTTTGTCTCAGAAACGATTTGACAGATGACGCAATTTTACATTC 926
Db      924521 GGTGATATGCTATTATTTATGAGCATCTCTGAATTTGCTGTAATGAAAAACATTTAATCTGT 924580
QY      927 TACAAGTTTCAGTCAGATGATGATGCTGAGAGACGCAAAAAGACTTGTCTAG---- 982
Db      924581 TTGAATTTAGAAACATGCTGTAATTTCTAAAGAAAGCTGATGAAATGCTTAAGACA 924640
QY      983 ---CCAAAACAGATGCAAGGGGTGATGCTTTTAAATGGGAAAAACATCTGAAT 1039
Db      924641 GATCCTGAAGCAAGCTGATGAGAAAAAGATTTTAAATTAAGATGATCTGAAT 924700
QY      1040 ACTCCAATTGACATTTGATACGCAAAAACAAGTTTAAAGCGATTCACAGATTTATTA 1099
Db      924701 ACTAAATGCGTCTTTATTTTGGC-AAAACAGTTTGAATGATTAATCTCAATTTTAA 924759
QY      1100 TGTATTAAATGCGATATGATGCTAGTGTGACAGCTCCACCTACAGTTGATGAATTTGA 1159
Db      924760 TGTATTAAAGGTAATGATGATTTAGTTGGGCTCGCCCAATTTGATTTGATGAGCTAGA 924819
QY      1160 AAAATATACCTCTGCGCAAAAAGACGATTTAGTTTAAACGAGGATTTACAGTCTCTG 1219
Db      924820 GCGTTATGAAGAAACGTAGATTATTACTTAATGCAAGACCGGATGACAGGCTTTG 924879
QY      1220 GCAGGTAGTGTGCTAGTAAATATCACAGACTTCGACGATTCGCTTGAATTTAGC 1279
Db      924880 GCAAGTCAGCGACGTAATTAATTC--GATTAATATCTCGGTTTATTTGATTTCTTG 924936
QY      1280 ATACATTTGATTAATTTGATGATCTGTCAGATATTTAAATTTTAAAGACAGTGAAGT 1339
Db      924937 GATATGTAATAAATTTGCTACTTGTGATGATGATTTTATTTTAAACAATGATGT 924996
QY      1340 TGTATTGTTGAGAGAGGAAAGTAAGTAAAGTATAT 1375
Db      924997 GGTTTTAAATCGTATGAGCTTATTAACGTTATTT 925032

```

```

RESULT 10
US-10-158-865-1
; Sequence 1, Application US/10158865
; Patent No. 684651
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fra

```

```
Patent No. 6846651
TITLE OF INVENTION: Thereof, and Uses Thereof
FILE REFERENCE: P186P2C1D1
CURRENT APPLICATION NUMBER: US/10/158,865
CURRENT FILING DATE: 2002-06-03 09:55:7,884
PRIORITY APPLICATION NUMBER: US 09/557,884
PRIORITY FILING DATE: 2000-04-25
PRIORITY APPLICATION NUMBER: US 08/476,102
PRIORITY FILING DATE: 1995-06-07
PRIORITY APPLICATION NUMBER: US 08/426,787
PRIORITY FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4416)..(4416)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
```



```
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...984
SEQUENCE DESCRIPTION: SEQ ID NO: 2527
US-09-107-532A-2527

Query Match
Best Local Similarity 51.4%; Pred. No. 6e-17;
Matches 384; Conservative 0; Mismatches 354; Indels 9; Gaps 1;

QY 4191 AAAAATGAATGATTTATTTACGTTATTTGACCAATTTAATAGTCCAAATTTCTTGA 4250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12 AAATAATAGTGTGAGATTGATTTATTTGTTCTGTGTATAAGTTGAAAAATATTTAG 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4251 TAAATGATTAACAGTTATTTATTAACCAATATTAATTTAGAGTTATTTCTGTAA 4310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 72 AAAATGTGATGATTCGATTTAGCCCAACATTTACTGATTTTGAAGTTATTTAGTGA 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4311 TGATGAGTACTGATGATTTCTGAGAAATTTGCTTAACTATATGAAAGATGAGAG 4370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132 TGATGATCTCCTGACATAGCGGAAATCTGTGATGATGCTGAAAAGATATATCG 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4371 AATTAATATTTACAAGAAATTAATGCGGCTGACAGATGCTGAAATTTGAGCTAGA 4430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 192 AGTACGTGTGATCCATTAAGAAACGCGGATTAAGCAGTGTAGAAATGCGGGATGGA 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4431 ACATGCAACGTAAATATTTGCTTTGTCGATTCGATGATCTATTTAGAATTGCAAT 4490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 252 TGTGGGAGAGGCAATTAATCTTGGCTTTGTTGACAGTATGATTAATATGATGAAT 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4491 GTTCGAGAGATCATGATTAATTAATCTGATTAATGCGATATATGACAGATGATTT 4550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 312 GTATGAAATTTTATAGAGATTTGAAAAATTCATGATGCGCAATTAAGCTCTGTGAAT 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4551 TTGTTAGTAGACGAAAAACGGTATACAAAGAAAAAAGAAATGTAATTTTCATGCTT 4610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 372 GATTCCTTTTATGTGATGATATTAAGAAAGCTAATTAAGAAAGAAATGTAATTTAT 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4611 AACGAGAGAGACATGTAATAAGATTTTGTGAGATCTAATATTAAGAAATTAATTTG 4670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 432 AATATAAAGAGCAATTAATCTGTTTGAAGAGAACGCAATTTTATGCAATACGATG 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4671 GTGCAAGCTTTATTCAGAGATTTATTAAGATTAATTAATTTCAATTTATTAATGAAG 4730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 492 GAATTAACCTTTATGTAAGACCTTTTAAAGATTAATCTTACTTGATGAGAAAAATTT 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4731 TATTTGTAAGATTTGCTTTTATTTGAGGCTTGAACATGTAAACAGTGTAGTAT 4790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 552 TGAAGATGCAATATATATATAGATCATTAATTT-----TCAACAGAAAAATTTGT 602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4791 TGATCTAGAGATTTATTTATTAATTAATGTCATTCGTAACAGTTCCGTTATTAATCAGAA 4850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 603 TGTTCCTCAATGAAGAGAAATTTCTATCTTCAAGAAATGATGATTAATGAGAAAG 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4851 ATTCTCTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 663 TTTTCTAATGAATTAATTTGATGTATTAAGAGCATGGCAATTAATTAACAAAGAGATTTT 722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4911 AAGAGAGTTTATGATTAATTTGATGC 4937
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 723 AGATGCTTTTCCAGATTTTGATGATTC 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-662-254B-27
; Sequence 27, Application US/09662254B
; Patent No. 693145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
```

```
FILE REFERENCE: UF-221C1XCI
CURRENT APPLICATION NUMBER: US/09/662,254B
CURRENT FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 09/086,651
PRIOR FILING DATE: 1998-05-29
PRIOR APPLICATION NUMBER: 60/224,479
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patent version 3.1
SEQ ID NO 27
LENGTH: 32392
TYPE: DNA
ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

Query Match
Best Local Similarity 43.2%; Pred. No. 1.1e-16;
Matches 1786; Conservative 0; Mismatches 2278; Indels 67; Gaps 20;

QY 2156 TTACAAGATTAATTAATTTTATTTATTAAGAAATATATGATTTGTTGAAAAATTTAT 2215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12462 TTTATTAATTAAGAAATTAATTAAGAAATATTTCTTAAATTAATTTGATTAATAT 12521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2216 GAAGTTCTAAGCAACCTAATTTACATCAAAATTAATTTTGTGAAAGATTAAAA 2275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12522 AAAAAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2276 CAATATGTTGAAAAATTTATGAGATCAAGAAATGAATTAATTAATTAATTAATTT 2335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12582 CATTAACGATCGAAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2336 GATTAATGCTTATCATTAATTTTCTCAGATTTTCTGAGAGGATTAAGATTAATCAT 2395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12642 TATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12701
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2396 CTTCCTCAGAGATGACACACCATTAATTTCTTCAATATCCTGATTAATTTTAA 2455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12702 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12761
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2456 TATTCAGAGATTTATTAATTTGATTTTCAAAAGATGAGAA--AATTAAGAAAAATG 2514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12762 TATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12821
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2515 GATTAATGAACGTTAAATTTAGATTAATTTCCATTAATTAATTAATTAATTAAT 2574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12822 ACAACATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2575 TATTAATGCTTTAGAAATTTTATTAAGATTAATTAATTAATTAATTAATTAATTA 2634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12882 TGTATTATCTTTAAA-----TATTAATAATTAATTAATTAATTAATTAATTA 12937
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2635 ATTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12938 ACCAAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12997
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2693 GGGTTTCCTCAGATTTTGTGCAATTTCTTTATCAATGAAGAAACAGCTT 2752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12998 CGTATTAAGAGAAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13057
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2753 ATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13058 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2813 ATGAATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13118 TTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2873 CAACATCTCTCTATTTGCTTAAAGATGATCTTAATTAATTAATTAATTAATTAATTA 2932
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13178 ATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2933 ATTTAGTTTATTTATTTGCTAGAAAGTTAAAAATTAAGAAATTAATTAATTAATTAAT 2992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Dh 13238 TATTCGATATATGTGTGTAAGGTAATATATATATAGAACATTATGATGATGCGAATCTA 13297
Qy 2993 TTAATACATAAAATATAAATAGTGTAT-TTGGAGAGATATGATATGTTAATATATTA 3051
Dh 13298 CATTTATATCTATATATAGATATATATATATATATGTTAGATGATCACAATATATAATATA 13357
Qy 3052 AATATGACCCGGAAATATTTTATTTTATAGTACTCTGTTGATATTTTATTTCCAGAC 3111
Dh 13358 TCTTTATATAGCAATCTTTTATTTAAATAATATATATAGAAAGAACCAATATATCTAAT 13417
Qy 3112 AAAAGATATATATTTTATTTAATTTTATGAAATTTAATTTTATTTCTATATATAATTTTGA 3171
Dh 13418 ATGATTTATGAATGTTATGATATATAGTTTA--TTAATCATTTAATTTAAATAATTTTGA 13474
Qy 3172 AATACATGCTAATATTTAAATAATGAAATTTTATGTTTATATGATGCTATATATATGTT 3231
Dh 13475 AAGGTTATATAATTTTAAATAATATATTTTATTTAATATAATTTATTTCTAATATTTGTTTA 13534
Qy 3232 TTGTTTCAGTA---GTCAAGATATGTTTGTGAAATTAATTTTGAAAGATTTATTTGCA 3287
Dh 13535 ATTCATACATATATTTGCGATATATAATATATACATATATAATAATATAATTTCTACATAT 13594
Qy 3288 GATTTTACTGCTCCCAATATTTGGATTTATGCAATATATGATATATATTTGATTTGATTT 3347
Dh 13595 ATATAATACATATGTTTACGAAATATAATTAATGTTAATGTAATTCAAAAATATCAAAATG 13654
Qy 3348 ATATAATATGATTTAATAAATAATTAATAATAGTATCTTTTATGTTTATGTTTATATTA 3407
Dh 13655 AAAATATATTTAATATTTAATAATAATTTATATGATATATACGAAACCGACGTTCTT 13714
Qy 3408 GGTATATCTGATGTATATATTTATCAAAATGGGAAAGATATGTTATTTTATGACAGAC 3467
Dh 13715 TATTAATCTTCCAGATATATATATATCAAAATTAATATTTGATTTAATAATATTTGGAAATATA 13774
Qy 3468 CTTATAGACTATGACTTATATACAGCGCTCAAAACAAAGTTGGTGGCTTTATGAAAC 3527
Dh 13775 TTTATATATATAGATTTTAAGTAATTTTATATATAATGATATATATAGACATATATAAATAG 13834
Qy 3528 TATCCATGTTAATACCACTACATATATAGTTTCATTCGTTAATCTTGACCTTATA 3587
Dh 13835 TATTAATTTGATATATGATATATAATAATATATATAATTAATTAATTTATTTATAAAC 13894
Qy 3588 AAAAATAAATAATGCAACAATTTTCTGTGCTTGTCTTTTATACGATCTATTTAAGT 3647
Dh 13895 TTTATCTTATTTATGATATATAATTTTATTAAGAAAGATATTAADAC--TCADATACATG 13951
Qy 3648 GGATGAGAAATGAGTTTATGCTATGCAATATTAATTAATGCTTTGTTATGAGATAT 3707
Dh 13952 GTAAGTATATTTGAATTTTATATATATACACAATCTATTAATATATATAACAATATATA 14011
Qy 3708 ATAGTGTGAAAAATTTGTTGATATAAAAACCTAATGTAATTTGTAATACCTATAT 3767
Dh 14012 ATATATCAAAATATATATCTGATATATA---AATAGAAATATCTACAAATTTGTAATGA 14067
Qy 3768 ATTTTAAATCTGAATTTGCTTACATGAAATTTTGCTTTTATTTATTTCTAGAGATCA 3827
Dh 14068 ATATAATATATCTATCGTTAGTATATAATATAAATAAATAAATAAATAAAGATTTCT 14127
Qy 3828 AGTACGAACTATGATTTATTTATCAAGAAATATGATATAAGTATTAAGAAACAT 3887
Dh 14128 AATATATATATATATTTTAAAAATTTAGAAAGCATCACACAGATATAAGATTTATAT 14187
Qy 3888 ATT---TTATTTGATATGAAATATCCGAATATTCATTTACGGAACTTGGCTCGGAAGT 3944
Dh 14188 ATTTTATTTATTTTAAATATATATCTACTATATGTAATTTATATATATATTTTGT 14247
Qy 3945 CATTCAGCTATATATCATTTTATTTATATAATCAGGAATAGTGGGATTTTACTGATG 4004
Dh 14248 TATTTTATTAATATTTTATTTATTTATTTGATTTTATTAATAATATATTAAGAGATTTGTT 14307
Qy 4005 TTTTCTTTTATTTATATGTTATATAAATAAGTTATGAGCTTAATGGGAAACGACATATTT 4064
Dh 14308 ATATAGTTTATTTTATATATATATATCTCTACATCATTTTCATTTATTAATAATATTTT 14367

Qy 4065 TATTTTACATCATTTAGCCATATTTTTCATATATGAAACAAATAGATCCGATTTATTTATA 4124
Dh 14368 CATATATTTTAAATATATATATATATGAAATATATATATATCATATATAGCG-TGATTAAGAT 14426
Qy 4125 TTAGTATATCTTTCTTTCTCAATAGCTATTTTGAATATATATAATTTTAAAAAGATATG 4184
Dh 14427 AAATATAATATATTTTATTTATGTAATGAGATGAGATATATTTATTTAAAAAAAATTT 14486
Qy 4185 GAGACAAAAATGAAATGATTTAATTTTCAGTTATTTGACAAATTTAATATGTCAGATTA 4244
Dh 14487 -----ATAATGATGATATATATATATATCATTTTCTATATATATATTTGT---TGATTA 14536
Qy 4245 TCTGATATAATGATTTAACAAGTATTTATTAACCAACATATACTAATTTTATGAGTTATCT 4304
Dh 14537 TCGTATATATTTGATCATCTTTCAATATATAGCATCTCAATATCTTCAAG---TTGAT 14592
Qy 4305 CGTAAATGATGAAATGATGATGATTTGAGAAATTTGCTTAACTATATGAGAAACGA 4364
Dh 14593 ATTTAATGTTTATTTCTTGTGCTTGTGTTTGTATTTGTTTAAATATTTATATATATC 14652
Qy 4365 TGAAGAAATTAATATATCAAGAAATTAATGCGCGTACAGATGCTCGAAATTTGCG 4424
Dh 14653 TTGCAATATATATTTTATATATTTTATATTTATTTAATTTAATGATTTGATATATCA 14712
Qy 4425 ACTAGAACATGACAGAGTAATATATGCTTTTGTGCAATCTGATGACTATATGAAAT 4484
Dh 14713 ACATATACCAATATGATATATATATTTATTTTATTTTATTTTATATATTTTGGTGG 14772
Qy 4485 TCGA-----ATGTTGAGAGAAATGCAATATATATACATGATATATATGCGATATAGCA 4539
Dh 14773 TATTCGAGTATTTACGAAATTAATATGAAATTTTATATATCTTAATAATTTTATATATC 14832
Qy 4540 GAGATGATTTTGTGTTAGTATAGCAAAAAAGGATATACAAAGAAAAAAGAAATAGTAT 4599
Dh 14833 CGTTAACAAATACATTTTTCAAATTTTAAATATGATATATACATATTTTATTTATTA 14892
Qy 4600 TTTCAATGCTTAAACAGAGAGAGAGACTGTAAGAAATTTTGTGAGATCTATATATGAA 4659
Dh 14893 TATTCGTCAACTCTTCATTTTAACTTTATATTTATTTATATATATTTTCAAAAAATTA 14952
Qy 4660 AATATATTTTGTGCAAGCTTTATTCAGAGATTTATTAAGATATATAAATTTCCAAAT 4719
Dh 14953 GAATATGATATATCTATTTATTTATTTATGAAATATGATCATTTATCTATATTTTCA 15012
Qy 4720 AATATATAGAAATTTGCTGAGATTTGCTTTTAAATTTGAGAGTCTTGAACATGATACA 4779
Dh 15013 TTTATTTAAATATTAACCTTTTGAATTTATCATATATAGTTCTATGAGCCATTTATTAAT 15072
Qy 4780 CGTGTATGATGATCTAGAGAAATTTATTTAATTTATGCAATTCGTAACAGTTGCTT 4839
Dh 15073 TTTATCGAATATATAAATCTAGGAAATATACGATTTTAAAAAATATATATTTTGTGATAT 15132
Qy 4840 ATTAATCAGAAATTTCTCTA-TAAATATATGATTTATAGTCACAAATTTGGAATTAACC 4898
Dh 15133 GTTTTTCATATATCTATATTTATTTATTTATGCGCATTTATGATATCTATATAATAT 15192
Qy 4899 CTTTATGTTTAAAAAGAGATTTAGTCATTTATTTTGTGCAAAAAAGTATTTAAAGAGAGGT 4958
Dh 15193 ATTTAATTAATTTGCAATATATAAATAATTTATTAATATCATTTTCAAAAAATATA 15252
Qy 4959 TAAATGTTTAAACAAATGATTTCAACAGATTTGTTGATATGATTTCTTGCCAAATAT 5018
Dh 15253 AATCATCAGATATATTTAAATAATTTCAATATTTTATTAATATGTTTGTGATATATATCAT 15312
Qy 5019 AGAGCTTATCGAAAAAGAAATGATGATATCATTTATTTAAAGCAAAAAAGATTTATTC 5078
Dh 15313 ACATACACATTTTAAATATATATATGATACATCTTTAAATCAATTTGGGTTTCTAT 15372
Qy 5079 AAAAAAGCAATTTTATGATGTTGATTTGATGAAATTTTGCCATAACTATATGTAATGTT 5138
Dh 15373 TACTAAATATATTTATTTATATATATATATATGATTTTAAATAATATCATTTTATTTATATAC 15432

QY 5139 ATATTAAGAAATTTTCAAAAGAGTAGAGTAAATAAGGATTAATTAAGTATTGTTTCCA 5198
 DB 15433 ATATATATGATTTATTTACCAATAATACATATATGAGATTTGTTATCGCTCAAAAGCAT 15492
 QY 5199 GTTTATATATGATTAATTAATTTAAGTAGTTGTTATAGAAAGCATTTATTAATCAAAATTA 5258
 DB 15493 CTTTAAATATCTAAATGTTAAATATATTTATTAATTTT---TTTATTTATTTTAAAGATTC 15548
 QY 5259 AAAATATAGAAATATATATGATAGATGAGCGCTGATAGATATCTGCTAAATATATGC 5318
 DB 15549 TTTTAAATCATTTATATGATTTGTTATATACATTAATTTATTAATTAATTTAATCAATCCACA 15608
 QY 5319 AAGGAATATGACAGAAAAGATTAATAAGATTAATTTTTCCTATCATCATAGTAGAGTA 5378
 DB 15609 TTAATATTCATCATGATTTTTCATCTAGTTATATATTTTGGTTGGTTGCTTACACA 15668
 QY 5379 TCAATATGCTAGAAATCATGGAATTAAGCGAGTACAGCTGAATATATTTATGTTGTTGAC 5438
 DB 15669 TTTCAATATATTAATTAATTTATTTATTCAGCGTTACTTATTAATTTCTTTTATTTATTTA 15728
 QY 5439 TCTGATGATGT-----TGTTGATGATGATTTAGTAGAAAATTAATATATTAATTA 5490
 DB 15729 TCTATGATATTTATTAATTAATCTATATATATTCGAAAATTAATTTTATCAAAATTTA 15788
 QY 5491 TAAATAGTAGAAGTATTTATCTGTTGTTGTAAGCTATCTTTTTCAGAAA---ATATTA 5547
 DB 15789 TAAATCTATCATATCTATTTTATTTATTTAATTTAGAAATTTTGGTTGATAGCTATGT 15848
 QY 5548 ATATATTTGAAAGTAAATCATCAAAATTTGATTTTGAAGCAATTAATCCGTGACAGACA 5607
 DB 15849 GATTTGTAATATTTTATTTACTAGACTGAATTCATATTAATTAATTTATATTTTTCAT 15908
 QY 5608 TGGAGAAAAAATTTTATGAAATTTGATATTAATTAATTTTCTACCTCGTTTGA 5667
 DB 15909 GGTATATCTAAACATTTATCATCAATTTTAAATATTAATTTTAAATTTAATTAAG 15968
 QY 5668 AACTATATTAAGAAAGTATCATACAGATCTTTTTCAGGAATCATGTTGAGGAAG 5727
 DB 15969 CGAATATTTTATTTAGAAATTTATTAATTTATTAATTAATTTGTA---TACACAAATA 16025
 QY 5728 ATTTACTTTTATCTGATTTATTTAAAGATTAAGATAGATAGTATTTAGCTGAAC 5787
 DB 16026 AAACATTTTATTCATGATATTTATCTATATTTATTAATTTAAACATTTATTTTGA 16085
 QY 5788 ATCTTATTTTATAGAGAGTATCTAATGTAAGTAATTTCTTTTAAAGAGTGTG 5847
 DB 16086 AAATTTTGTATTTATCTATTTTAAATTAATTTATTTATTTATTTATTTATTTATTTATTT 16145
 QY 5848 TTTTGAATTTGAAATTTTGCAAAAACAGTATGATTTGTTAAGCAATATATGCTG 5907
 DB 16146 TTTTGTATTAATATTTATTAATTAATTTCTATTTGCGAATCAACGTAATTTATTTA 16205
 QY 5908 AGGATTTTGCATCATCAATTTGTAAGATCTATAGCTGGAAGATTTTATTTATAGCT 5967
 DB 16206 TTGATTTATTTCTAATA---TACATCTCTATCTAATGCTAATAATTTGATTTAAACCT 16262
 QY 5968 TACTTAATGTTTAAATAAGAAAACAGTCTATTTTGAACAAATTTTATTTTGAATC 6027
 DB 16263 TGATCTAAATATACATTTTAGACATTTAGCAACATATATACCATTTCTGCAATTTTAAATTA 16322
 QY 6028 TTTTATTAATTAATTTTATTTTAACTTTGTTAAAGTATCTAACAATAATTTCTTCTCTA 6087
 DB 16323 TTTAAATGATTTGAATTTGTTGTTATAGAAACATTTTGTACATATATGTAAGATTTG 16382
 QY 6088 ATTTTGTATTAAGATTTGTTTGCACAAAGTTTAAATAATTTATGTTATTAATGAG 6147
 DB 16383 CAAGTTAAACATCTATCATTTTACTTTAATTTATTTATTTCTCTACATATTTGACAAAT 16442
 QY 6148 AAGATATCATGATATCTATTTAGTAAATTTCTAATTTGACTATATATATATGTAAGAA 6207
 DB 16443 TTTCTATTTTAAATAATCATTAATATATATTAATTAATTTTATTTTATTTTATTTT 16502
 QY 6208 AATATTTATCTAATATGATATGATAGCATTTGTAATATGAGACTTCAAAACATA 6258

DB 16503 ATATTTATTTTCAAAAAAATATTAATTAATTAATTTATTAATTTGAAAAAATATTAATA 16553
 RESULT 13
 US-09-662-254B-26/c
 ; Sequence 26, Application US/09662254B
 ; Patent No. 693145
 ; GENERAL INFORMATION:
 ; APPLICANT: Moyet, Richard W.
 ; APPLICANT: Li, Yi
 ; APPLICANT: Bawden, Alison Louise
 ; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
 ; FILE REFERENCE: UP-221C1XCI
 ; CURRENT APPLICATION NUMBER: US/09/662,254B
 ; CURRENT FILING DATE: 2000-09-14
 ; PRIOR APPLICATION NUMBER: 09/086,651
 ; PRIOR FILING DATE: 1998-05-29
 ; PRIOR APPLICATION NUMBER: 60/224,479
 ; PRIOR FILING DATE: 2000-08-10
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 26
 ; LENGTH: 50000
 ; TYPE: DNA
 ; ORGANISM: *Ambactra moorei* entomopoxvirus
 US-09-662-254B-26
 Query Match 2.1%; Score 145.8; DB 3; Length 50000;
 Best Local Similarity 42.8%; Pred. No. 1.3e-15;
 Matches 1879; Conservative 0; Mismatches 2432; Indels 84; Gaps 19;
 QY 1865 CGATTGATTAAGAGATTTGATTTATTTGAAAAAATGAAAGTAAACCGAATATTT 1924
 DB 27465 CGATTGATTAAGAGATTTGATTTGAGATTAAGAAAAAGAAATTTATATATATATATAC 27406
 QY 1925 ATTCGAACGAGATTTCTGATATATTCAGATATTTGCAATATTTGCAATTTTCTCAGT 1984
 DB 27405 ATATTTATTTAGATTTATTTATTTAGTAATTTAAATTTATTAAGAACCATATATTTAT 27346
 QY 1985 TACAAAGAAATGGAACATATATTTAACAATCGAAGTATTTGCCACGAGGCCCC 2044
 DB 27345 TAAACATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 27286
 QY 2045 GCTACTTTATGAAATTCATTTATTCAGAAAGAAAAAACAATTTGTTCTTACAGAAAA 2104
 DB 27285 TAAATTAATATATGATGATGATTTCTATTAATTTGAATGAATTAAGAAAAAATACTAA 27226
 QY 2105 AAGTAT-GGTGAACATGTAATGATCATCAAGTAGGTTGTAAGAAATTTTACAAGA 2163
 DB 27225 TATTTCTAGATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 27166
 QY 2164 TATATATTTTATTTATTAAGAAATATATGATGATTTGTTGAATAATTTATTTGAAGTTTC 2223
 DB 27165 TGAATTAATTTATTTATTAATTTATTAATTTATTAATTTATTTATTTATTTATTTAT 27106
 QY 2224 TAAGCAACTAATTTTATCATCAATTAATTAATTTTGTGGAAGATTTTAAACAAATAGT 2283
 DB 27105 TGAGTGAAAAAATAGTAACACCTTAATTAATTAATTAATTAATTAATTAATTAATTA 27046
 QY 2284 TGAATAATTTATTTAGAGATCAAGAAATGAAATTAATTAATTAATTAATTTGATTAATG 2343
 DB 27045 TTAATAATATCTAAACTATGATTTGAATTTATTTTATTTATTTTATTTATTTATTA 26989
 QY 2344 CTATCATATTTTCTCAGATTTTATCTGAGAGGATTCAGATATTTATCATCTTCTCTC 2403
 DB 26988 ATTTATATATTTAAATTCGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 26929
 QY 2404 AGGAGATGACACCATTTAGTTCTTGAGAAATCTGTAATTTATTTTAAATTTCTCA 2463
 DB 26928 ATTTATCTAATTTATGAAAGATTTTGAAGATA---ATGAAGATGATATTTATTTCTACT 26874

QY	2646	GGATTATATGTTGAATTATCAAAAGATGACGAAAAATATAAGAAAAATGATATATGCA	2523
Db	26873	TAGATCTACTTAAAAAAAATATTTAATGATGCTGATATGAAAAAGAAATTAATTAATAT	26814
QY	2524	ACGAGTTAAATGTTAGCAGATTATTTCCCTAATATATCGAAAAAAACATATGATATGCT	2583
Db	26813	AGAAAAATACACATTTATGATGATATCACTTTATATACGTATATATTAATTAATGATATGA	26754
QY	2584	GTTTGAATTTTATTAAGAAATGATCGAGCTTTTGAATCTATTTTCAAGAATGTTGTT	2643
Db	26753	ACTAGGTTCTTGTTTAAAGATATATGTTCCGTTATTTAATATGAAAAAATGTTGAT	26694
QY	2644	TATGTATGAAATAAAAACATGGCTATAGAAATAGATTTGGTTCTAATTTGGTTCCGCTT	2703
Db	26693	AACATAAGAAATTTTAAATTTTATTAATATTTCCAAAGTGG-----GGAAGTGTTC	26442
QY	2704	CCACATGATTTTGGCAATTCCTTTATCAATGAAGAAAGCAACGCTTATTTATTTAG	2763
Db	26641	CCCATATTAATTTAGCAATTAATATACAAATATATGAGAAAGATATATGATCATTTTAAA	26582
QY	2764	TAACTCTAAATGTCAGATGACATTAATTAACAGCAATTAAGAAAAATATGATTTTCA	2823
Db	26581	AAGAAATTAATATGAAATTAATATATATATTAATATATATATATTAATATAATAT---	26525
QY	2824	AATGATTAATCTAATATATGGAATTTAGATATATTAAGTGAAGAAAAATCAACATCTCT	2883
Db	26524	AATATATTTAAAAAGTGAATTTAAATTTATGATATAGCATGCTTATGTTCAAAAAACATGG	26465
QY	2884	CCATATGCTTTACAGATGATCTCATATGATGAAATGCTAATGCAAGAAATTTAGGTTT	2943
Db	26464	TTAATATATATACGTTGATGATTTCAAGAAATCAATTAATGATGATGAACAACATTAATCT	26405
QY	2944	TTATTTGCTGAAGATTTAAAAATGAAATTAATCTAAATTTTAAAGAAATTAATCTAAA	3003
Db	26404	AAAAATGATGAATAATATATATATAAAAATGAAAAAATATATTTGAATATTTCTACTTAT	26345
QY	3004	AAATPAAAATGTTGATTTTGTGAGATATGATGTTTAAATTAATTAATATATGACCGG	3063
Db	26344	ATATCAAAATATTAATTAATAAAGA--AATAGAAGATATTAATAAATTAAGAAATTTTAATAAT	26286
QY	3064	AATATTTTATTTTAAAGTCTCTGTTGATTAATTTTATCCAGACCAAAAGTATGAT	3123
Db	26285	AAAAAGAAATAATATATATTAACAAAGATTTAATATGAATATGAAAAATTTTATTAATTA	26226
QY	3124	TTTATTAATTTTATGAAATTTAATTTTATTTCTCATATPAATTTTGAAGAACTATAGCTTA	3183
Db	26225	TGTTAAAGAAAAAATAAAAATGTTATATTTATATAGCTACAGTCAAGCTTATGCTAT	26166
QY	3184	TATTPAAAAATGAATTTTATGTTTATTTATATGCTATATATATATGTTTGTTCGATAG	3243
Db	26165	GATTAATATATTTTAAAGATGTT-----AAACATGATTAATTTCTCTAGACA	26118
QY	3244	TCACAAGTATGTTGTTGAATAATTTTGAAGATTAATTGCAATTTTACGCTCCCA	3303
Db	26117	ATCAAAATTTTATCTTCACATTAACACTGGAAGATTTTATATATATGTTATATGAAAA	26058
QY	3304	TAAATTTGATTTTGAATATGATATATATTAATTTGATTCATTTAATPAATATGATTTATA	3363
Db	26057	AGTATTTATATATCTTAACACAAAAATTAATTCATGATCTTTTATGATTAAT--TTAAG	26000
QY	3364	AAAAATTAATAAATAGTATCTTTTATGATTTTATGATTTTATAGATATATCGATTTGT	3423
Db	25999	GATAAAAAGAAAAAATTTTGTATACATTAACAATATTTATTAATAGTATGTTAAT	25948
QY	3424	ATATTTATCAAAATGGGAAAGATATTTGATTTTATGACAGACACTTATATGGACTTAGCT	3483
Db	25939	TTAGTAATTAATAAATAATTAATGAACATATGATTAATTAATTAATTAATTAATAAACA	25880
QY	3484	ATCTTATTAACAGGGGTCAAAACAAGGTGTTGTTGCTTATGAACATATCTAGCTTAATAT	3543
Db	25879	TTATCTGAATATGATTAATCTTAAACCTATTAATACCAATTAATAATTAATCAATTAATATGA	25822
QY	3544	CCACATCAATATATAGTTCAATTCCTGTTAATCTTTGCACTTAATAAAAATTAATAATCAAC	3603

D	25819	GAAATGCAATTAAA--TGAAATTTAAAGAAAAAATACATATATTTCTAGATGAATATATTA	257653
Q	3604	AAATTTTTTCTGTGCTCTGCTTTTATACGATCTATTTAAGTGATCGAAATTTGTA	3663
D	25761	AAATATATATTTAAAAATATCTGAAATN--GAATATTTAAATTAATTAATATGATTA	25704
Q	3664	GTTTATCGTAGCAATATTAATTAATATGCTGTGTATGAGATATATAGTGGAAAAATTGG	3723
D	25703	TATTTAACAAAAACAATTATGGAATTAATTTATGAGCAAAATTTTGGTGAAAAAATGTTA	25644
Q	3724	CTTGAATAAAAAGCTAATGATTAATTTGTATATCTACTTAATTTTAAATGTAAT	3783
D	25643	AACACCTAATAAATTATATGATGATAAAAAATTAACAATTAATGATTAATATGTTTTT	25584
Q	3784	TGCTTTACCATGAAATTTTGGCTGTTAT-----AATTCGAGATCAAGTAAACA	3835
D	25583	TTCAATATGCTAACATTTTATGATATTTTTTCTAACTTAATTAATCTATCATATATTTAT	25524
Q	3836	AGCTAGATTTATTTATTAACAAGAAATATGATTAAGATTAAGAAACAATATTTTAT	3895
D	25523	TTTGGTTCTAATTAATTTCTAATTAATTTGCAATTTTATTTTAAATGTCGAATATCGC	25464
Q	3896	TGGATATGCAATATCCGAAATATTCGTTACGGAACTTGGCTCGGAAGCATTCAGCGTA	3955
D	25463	ATCAAAAAATTTATTAAGATATATATCTATATTAATCTGTAATATTTAAATTAATATAT	25404
Q	3956	TATATCATTTTATTTTAAATTCAGAAATAGTGGGTGATTTTATACGATGTTTCTTTTT	4015
D	25403	CATATTTATATTTTAAATCTGTATTAATACCATGTATTAATGCAATCGTCGTGATAT	25344
Q	4016	TTATGTTATAAAAAAGTTATGAGTTATAGGGAAACAGACATATTTTATTTTATCATC	4075
D	25343	TTCAATAAATAATTAATTAATTAATTCAGATATATTTATTAATATCATTAATCCATTTCTC	25284
Q	4076	ATTAGCAATTTTTCATATATGAACAATAGATCCGATTAATTTATTTATTTAGTACTAT	4135
D	25283	TATTCAAATACCAATTAATGATTTTATATCATATGAATTTAAATTTTATTTTATTAAT	25224
Q	4136	CTTTCTCAATAGATTTGGAATTAATTAATTTTAAAGATATGAGACAAAAA	4195
D	25223	TATTAATTTATATCTGTAT-----ATGATTAATGATTAATTTGATATATCATTA	25176
Q	4196	TGAATGATTTAATTCAGTTATTTGTAACAATTTATATGTCACAGATTAATCTGTATTAAT	4255
D	25169	TATTTATTCGTTAATATATATTTATTCATTTTAAATTAATTAATTAATTTTCAAA	25110
Q	4256	GTAATTAACGATTTATTAACCAACATATATCTAATTTAGAGTTATTTCTGTAAATGATG	4315
D	25109	TTATTTAGTTTATTAATTAATTAATTAATTAATGACATATATTAATGATTAATAT	25050
Q	4316	GAGTATCTGATGTTCTGAAAAATTTGTTAAACATATGAAAGAACATGGAAGATTA	4375
D	25049	GTAATTTCTTTTAACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	24990
Q	4376	AATATTAACAAGAAATTAATGCGCGTCTAGACAGATCTCGAAATTTCCGACTAGAACATG	4435
D	24989	GTTATTAACATCTGTCAAAAAGAAATTAACAACAATTAAGTGCTTCAAAACAATTAACATTA	24930
Q	4436	CAACAGGTAAATATATGCTTTTGTGCAATCTGATGATATATAGAAATTTGCAATGTTG	4495
D	24929	TACATTTGTAATATGTTTTTAAAGAAATATCAAAATCCAGTATTTTCATATGATATGATAG	24870
Q	4496	AGGAATGATGATTAATATATATCTGAGATTAATGCCATATAGCAGAGATATGATTTTGT	4555
D	24869	ATGATTTTCTAAAACTAATCAATGATTAATATTTAAATCAAAAGATTAACAATCGCGT	24810
Q	4556	TAGTAAAGCAAAACGGTATTAACA---GAAAAAAGAAATAGTATTTTCAATGCTTTAA	4612
D	24809	TAGATTTCTGATATTTTAAATTAACAAGATTAATTAATTAATTAATTAATTAATTAATTA	24755
Q	4613	CGAGAGAAAGACTGTAAAAAGATTTTGTGCAAGATCTAATATAGAAAAATTAATGTTTGGT	4672

Db	24749	TTAATATGCTCTAATCCGATNGGTAAATATATGTAATATTAACTAAAGACGATATCTTAGAA	24690
OY	4673	GCAAGCTTATTCACGAGATATTATAAAGATATAAAATTCGAATTAATATGAGTA	4732
Db	24689	TTAATATATAATTTTGGAGATTTTACAAATTTATATATAAAAAATATATACAAAACTA	24630
OY	4723	TTGGGAGGATTTGGCTTTTAAATTGGAGGCTTGAAACAATGTAAACCGTGTAGTGGT	4792
Db	24629	TATCTTACGAATTCGTTTAGAT-----AGAAATTTGATATTTCTTAATATTA	24582
OY	4793	ATACTAGAAATAATTAATATATATATGTCATTCGTAAACGTTGCTTAATTAATCAGAAAT	4852
Db	24581	ATTATGTATAATTAATTTCAATTATGCAAACTAATATAAAGACGATTAACCGAAATAT	24522
OY	4853	TCTCTATATAATATATGATTTAGTCAACAAGTTGAGAAATTACCCCTTAAGTAAAAA	4912
Db	24521	GGAATTAATATATATATGGAATTATATATATAAATCATGATTAATCATTAATATAAATTA	24462
OY	4913	GAGAGTTAGCATTAATTTTGAATCCAAAGCTATATAAGAGAGTTAAATGTTTAAACA	4972
Db	24461	TATTTTGAACCAATTAATATATATTTTATGAAATATATAAAAAATTAATATATGTTCCGTA	24402
OY	4973	AAATGTATTCACAGATTTGTTGGATATAGAGTTCTTGCAATATAGAGTCTTATCGAA	5032
Db	24401	ATATTTATTTAATAATATGTTATTAATAATTT-----GTTGATGTTTTTCTATATA	24350
OY	5033	AAGAAATACGTAGATATCCATTTATATAAGCGAAAGATATTTATCAAGAAAGCATTTAG	5092
Db	24349	AAAAATATTAATAAATAATATATATATATATTAATTAATTAATAAATATGTTATATA	24290
OY	5093	TTAGCTGTATTTGATGAAATTTTCGCTAACTATATGTAAATGTTATATAGAAATTTTC	5152
Db	24289	CTAATATATTAATCTGAGTTGACTAATGATTTTATATATTAATCTATGATATAGTTGATGAT	24230
OY	5153	AAAAGCAGTAGAGTAAATAATGATATAAATTAGTGTAATGTTCCAGTTTATAATGTAGA	5212
Db	24229	AAATCATCATATAATATTTTATTTATATAATGTCTATATGATCTTAAGAGTGTAAATA	24170
OY	5213	TAAATATTTTATAGTATGTTATAGAAACATTAATTAATCAAAATATATAAATAATAGAAAT	5272
Db	24169	GTTATATCTATATTTGTTTCCAATTATATATTTTATATAATATATAATATATATATATTA	24111
OY	5273	ATTATGTATAGATAGCGCTCTGTAGATGATTCTGCTAAATATATGCAAGGATATGCAGA	5332
Db	24110	AGATTTTATTTTATATGATATGATATATGATTAATAATATATATACAAATGATTCGATCA	24051
OY	5333	AAAAGATTAAGAGTAAATAATTTTTCACATATCATATGAGAGTATCAAAATGCTAGAAA	5392
Db	24050	ATGCTGATATTTGTTTAAAAATTTTATCTGATGCTGTGCACGAAAGAAATTAATAATGT	23991
OY	5393	TCAATGGAATAAAGCGGAGTACAGCTGATATATAATATGTTTGTGACTGTAGATGTT--	5450
Db	23990	TAAATGATCATATCGTAATATCAATTTTGAATTTTAAAAAATCAATATGTAATATATAA	23931
OY	5451	GTTGATATGATAGTATGTAAGAAAAATTATTTTATATTTATATAAATAATGTAAGTGAATTA	5510
Db	23930	TTTATATATTTGGATTAACGAAGTAAGTCTAAATTTATATCTAAATATATTCACATATATTA	23871
OY	5511	TCGTGTTTGTGATCGCTATTTTCAGAAATATATAATTAATTTTGAAGTAATATCA	5570
Db	23870	ATTTTAAATGTTATTAAGACATTAATTAACAAACAATGATTAATATATATTTGCAAAATA	23811
OY	5571	AATATATGATTTTGAAGCAATTAATACCGTGCAGACAATGGGAGAAAAAATTTTATGAAT	5630
Db	23810	TATTTAATATATATATTTACTTTTTTTTAACTTTTGGTATGTTGTAGAGATTAATATAT	23751
OY	5631	TTGTATATAAATATAATTTTTTCTACTCTGTTTGTAACTATATTAAGAAAAAGATATCAT	5690
Db	23750	TTTCAATATCATCATGATATATATATATGTTGTATATCTTGTAAATACATCTTCGCAT	23691
OY	5691	ACAGATCTTTTCAAGGAATCAATGGTATGAGGAAGATTTACTTTTATCTGATAT	5750
Db	23690	GTAATAGTTTCTTCAATATTTATTTATATATTTTAAACGGGTGACTATATATATCTTTATAT	23631

QY	5751	TTAAGAATATAGAGTAGAGTTAG - TATTTGACTGAAATCTTATATTTTATAGAGAG	5809
Db	23630	TTTTGGTATATCAATAGTGGTCTTCTATTTTAAACGTGAAAAATTTGCATATTTAA	2357
QY	5810	TATACTAGTACAGTAAATCTTTTAAAGAGGTGTGTTTTGCATTTGAAAAATTTGCA	5869
Db	23570	TTAATTTTACTCTCTCTTTTATTTTATTTCTATTAACATTAGATCATTTGATTAATGTAT	2351
QY	5870	AAAACAAGTACAGTATTTGTTTAGCAATATATNGTGAGATTTTGACGATCAATGT	5929
Db	23510	ATATATAATTATTAATTAATTTTCCAAATATTTTCCAACTGACTTATTTTAAAGTAA	2345
QY	5930	TAAAGATCTATAGCTTGGCAAGATTTTATTTATAGCTTCTAAATGTTTAAATACGGAA	5989
Db	23450	TCAATTTTACTCCGCAATATTAATTCATTTTATTTATNGATGTATTCACATATTTTAA	2339
QY	5990	ACAGCTATTTTGTGACAAATTTTAAATTTTATAGAAATCTTATATAAAAAATATATTTAA	6049
Db	23390	ATAAATATATTTTTCATCATCTGATGTTTGTATTAATAAATATATGATTTGTATTTGAA	2333
QY	6050	CTTGTTAAAA---GATCTPACAAAAATCTTTGTCTAAAAATTTTGTATTAAGATTG	6105
Db	23330	AATATATAAAATTCGATTAATATTCACATATTTAGATGTTTATATATATATTTGAATGTTG	2327
QY	6106	TTTGGAACAAAGTTTTTAAAAAAATATATAGGTATATATAGAAATATCATGATCTACTA	6165
Db	23270	TTTTTTATTAACGATTAATAGTTTATATAAATGTGCATTTTAAAAATGAAATCATATATATTA	2321
QY	6166	TTAGTAAAAATTTCTATATATGTACCTATATATAATATAGTAAAAAAATTTATCTAAATGTA	6225
Db	23210	TGATAGACGGATTTTATATATTTTGGATGATPAAATATATGAAAAATCTATATTTAATTA	2315
QY	6226	TAGATAGCATTTGTA 6240	
Db	23150	ATATATTATTTATATCA 23136	
RESULT 14			
US-09-134-000C-2987			
; Sequence 2987, Application US/09134000C			
; Patent No. 6617156			
; GENERAL INFORMATION:			
; APPLICANT: Lynn Doucette-Stamm et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO			
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: 032796-032			
; CURRENT APPLICATION NUMBER: US/09/134,000C			
; CURRENT FILING DATE: 1998-08-13			
; PRIOR APPLICATION NUMBER: US 60/055,778			
; PRIOR FILING DATE: 1997-08-15			
; NUMBER OF SEQ ID NOS: 6812			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 2987			
; LENGTH: 993			
; TYPE: DNA			
; ORGANISM: Enterococcus faecalis			
US-09-134-000C-2987			
QY	Query Match	2.1%; Score 145.4; DB 3; Length 993;	
	Best Local Similarity	52.0%; Pred. No. 9.8e-16;	
	Matches 382; Conservative	0; Mismatches 341; Indels 12; Gaps 2;	
Db	6164	TATATAGTAAATTTCTATTAATGTGACCTATATATAATGTAGAAAAATATTTATCTAATG	6223
	18	TATGCCAAATTTAGTATTTATTTGCTCTGTAACATGTGAAAAATATTTGAAAAATG	77
QY	6224	TATATAGTACATTTGTAAATCAGACTTACAAACATATATAGATTTCTTGTGTAATGACGG	6283
Db	78	TGTACGCTTATTTTACTCAAAAGTTTACTGACTTTGAATTAATTTCTGTGTGACGATGG	137
QY	6284	TATGACGATATTTCCGAAGAAATTTCTTTTATGATATAGCAAGAAAGATAGTGCATTCG	6343


```

Db      138 CTCTCCAGACAGTCTTGAGCAATGTGTGATCAGTTTGCTGAACAAGATCAACGGGTTAA 197
QY      6344 TTATTTTAAAAAGAAACGGGGGGTATCAGATGCCGTAATTTATGCGATAAGTCGCGC 6403
Db      198 AGTTATCCCTAAAGAAATGTTGGCTAAGCCAGATGCTCTTAATGCTGGAATTTGAAATGAC 257
QY      6404 CAAGGTGACTACTTAGCTTTTATAGACTCAGATGATTTTATTCATTTGCGAGTTGATCCA 6463
Db      258 AACAGGTGATTTTAGTTTGTGTAGATAGTATGATTCATTTGAGATGATATGATGA 317
QY      6464 AGCTTTACAGAACCAATTTAGAGAGAGATGCCCTTTGGGAGTTGCTGTGTTATGATNG 6523
Db      318 ACTATTTATACAAATATCTGTAAGAAAGATGCCGATTTATCAATCTGTGTATTTATGA 377
QY      6524 GGTATGCTTGGGGGCACTTTCTTAACAGACAGCGCTTCCACAAATCAGGCTGTTC 6583
Db      378 TGTCTATG---AGGGGAAAGAGCCCAATTTGTAAGATTTAATACGGGAATTTTCCA 433
QY      6584 GAGCGGACAGAAATGTTGTAAAAAGCTGCTAGAGCGGATGTCTATCGCTTTGTGTGCGC 6643
Db      434 GAGAGAAACATTTATTTGTTAA-----TTTACAAAGAAATATTTATCTCGTACATGC 485
QY      6644 CTGTAAATAACTCTATPAAAAAGAACTATTTGAAGATTTTCATTTGAAAAGGTAAAT 6703
Db      486 TGTAAATAAGTTATATAAAGAAACCTTTGGCGATTTAAGATCTTAAAGAAAGTA 545
QY      6704 TCATGAATGAAATCTTCACTATGCTTCTCTATGAGTTAGAAAAGTTGCAATAGT 6763
Db      546 TCATGAAGATCTTATTTATAGTATGATTTGTTGTTCTGAATGTCAGAAAGTCTTATTTGA 605
QY      6764 TAAGAGTCTTGTACTATTTATGTTGACCGAGAAATAGTATCAAACTTTAGCATGAC 6823
Db      606 TTCAACAAAAATATATCATTTATCATGGAATGGGACATTAATACAGAAATTTTC 665
QY      6824 TGACCATCGCTTCATCTGCTACTGGAATTTCAAAATGAACGATGACTTCTATGAAG 6883
Db      666 AGATAAACATTTGAGTTTATAGAGCTTGGGAGAAATGAATTAATTTGAAGTTAA 725
QY      6884 TAGAGAGATTAAGA 6898
Db      726 GGGAGCAGTTATAGA 740

```

RESULT 15

```

US-08-961-527-71
; Sequence 71, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373

```

```

; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-527-71
;
Query Match      2.1%; Score 144.6; DB 3; Length 32768;
Best Local Similarity 52.3%; Pred. No. 2e-15;
Matches 368; Conservative 0; Mismatches 329; Indels 6; Gaps 2;

```

```

Search completed: December 25, 2005, 16:25:09
Job time : 816 secs

```


Db 181 TAGAGTTGAAAAACATTTAACTATAGTATATAATTTGCAATTTTCTTACGCGAGTAT 240
Qy 241 CATTTTGTGGAGATATATTTGCACTTCAAGAGTGTGCGGTATTTCACTTAA 300
Db 241 CATTTTGTGGAGATATATTTGCACTTCAAGAGTGTGCGGTATTTCACTTAA 300
Qy 301 TAAACTGCTTTTGTATACCTATTTACGTATTTTAAAGAGTTAAGATAGCTTTC 360
Db 301 TAAACTGCTTTTGTATACCTATTTACGTATTTTAAAGAGTTAAGATAGCTTTC 360
Qy 361 TATTTTCCAGCATCTATCAAAAAAGCATTTCTAATTCAACGCGTGAACATGGGAAA 420
Db 361 TATTTTCCAGCATCTATCAAAAAAGCATTTCTAATTCAACGCGTGAACATGGGAAA 420
Qy 421 ATATGCAAGTTTATTTGAATCAGATAAACAATTCAAAAAAATCTTGTGCAATTGTAG 480
Db 421 ATATGCAAGTTTATTTGAATCAGATAAACAATTCAAAAAAATCTTGTGCAATTGTAG 480
Qy 481 TTTTAGGTACAGAAATAGATATAAATTAATTTATCAATTACCGCTCTATTAATCTGTGAAG 540
Db 481 TTTTAGGTACAGAAATAGATATAAATTAATTTATCAATTACCGCTCTATTAATCTGTGAAG 540
Qy 541 AAGCTATAGATTTTCAACAGGGAAGTGTGCAACAGCTCTTAAATATCTACCAAGT 600
Db 541 AAGCTATAGATTTTCAACAGGGAAGTGTGCAACAGCTCTTAAATATCTACCAAGT 600
Qy 601 AGTTTATAGATTTTCAACAGGGAAGTGTGCAACAGCTCTTAAATATCTACCAAGT 660
Db 601 AGTTTATAGATTTTCAACAGGGAAGTGTGCAACAGCTCTTAAATATCTACCAAGT 660
Qy 661 TTTGATTTAATTCATTCGTTTCTGCGTTGAAAAAACAATAATCCAACTGCTAGGTG 720
Db 661 TTTGATTTAATTCATTCGTTTCTGCGTTGAAAAAACAATAATCCAACTGCTAGGTG 720
Qy 721 ACCATGCAATTTTCACTTTTCCAAATTTTATAAGCTACTCATATCATATGAAC 780
Db 721 ACCATGCAATTTTCACTTTTCCAAATTTTATAAGCTACTCATATCATATGAAC 780
Qy 781 GACTTTTGAATATACCTGAGCGGTGTGCGGTAAATTTATGAGTATGTTCTATTT 840
Db 781 GACTTTTGAATATACCTGAGCGGTGTGCGGTAAATTTATGAGTATGTTCTATTT 840
Qy 841 TGTTAATTCATTTATTCGTAGAGATGTGTGCAACGCGTATTTTGTGCTCAAGAA 900
Db 841 TGTTAATTCATTTATTCGTAGAGATGTGTGCAACGCGTATTTTGTGCTCAAGAA 900
Qy 901 GACAGAAATGAGCCATTTTACATTTCAAAATTTTCATGATGTATGTGATGTGAGG 960
Db 901 GACAGAAATGAGCCATTTTACATTTCAAAATTTTCATGATGTATGTGATGTGAGG 960
Qy 961 AGGCAAAAAAGACTTGTGAGCAAAACAGATGCAAGGATGAGTATGTTTAAATGG 1020
Db 961 AGGCAAAAAAGACTTGTGAGCAAAACAGATGCAAGGATGAGTATGTTTAAATGG 1020
Qy 1021 GAAAAAGCATCTAGATTTCTCCAAATTTGCAATTTCTATACGCAAAAAACAAGTTTAAAG 1080
Db 1021 GAAAAAGCATCTAGATTTCTCCAAATTTGCAATTTCTATACGCAAAAAACAAGTTTAAAG 1080
Qy 1081 AGTTACCAAGTTTATATATGTTTAAATGTCGATATGAGTCTATGTTGTACAGCTCAC 1140
Db 1081 AGTTACCAAGTTTATATATGTTTAAATGTCGATATGAGTCTATGTTGTACAGCTCAC 1140
Qy 1141 CTACAGTTGATGATTTGAAAAATATCTCTGCTCAAAAAGAGCGATGAGTTTAAAC 1200
Db 1141 CTACAGTTGATGATTTGAAAAATATCTCTGCTCAAAAAGAGCGATGAGTTTAAAC 1200
Qy 1201 CAGGATTTACAGGCTCTGCGAGGTTAGTGTGCTAGTAAATATCAAGACTTTCAGCAGC 1260
Db 1201 CAGGATTTACAGGCTCTGCGAGGTTAGTGTGCTAGTAAATATCAAGACTTTCAGCAGC 1260
Qy 1261 TAGTTGCGTTGCACTTATGCAATCATGTAATTTGAGCTATCTGTGATGATTTAAATTT 1320
Db 1261 TAGTTGCGTTGCACTTATGCAATCATGTAATTTGAGCTATCTGTGATGATTTAAATTT 1320

Qy 1321 TATTAAGACATGAAAGTTGATTTTGTGAGAGGGAAGTAAAGTAAATATGAAG 1380
Db 1321 TATTAAGACATGAAAGTTGATTTTGTGAGAGGGAAGTAAAGTAAATATGAAG 1380
Qy 1381 TTTGTTGTGCGGTTCTCAGGGGGAATTTGACTCATCTTGTATTTGTTAAACCGTTTT 1440
Db 1381 TTTGTTGTGCGGTTCTCAGGGGGAATTTGACTCATCTTGTATTTGTTAAACCGTTTT 1440
Qy 1441 GGAAGGAAGAAACGTTTTTGGGTACATTTGATTAAGAGATGCAAGAAATCTTTTGA 1500
Db 1441 GGAAGGAAGAAACGTTTTTGGGTACATTTGATTAAGAGATGCAAGAAATCTTTTGA 1500
Qy 1501 AGAATGAAAAATGATTCATGTTTCTTCCAAACAATTCGCAATCTATTAATTTAGCA 1560
Db 1501 AGAATGAAAAATGATTCATGTTTCTTCCAAACAATTCGCAATCTATTAATTTAGCA 1560
Qy 1561 AAAAATCTTCTTACCTTTCAAAATTTTACGATGAGAAACCAGATGTTTATTTTAT 1620
Db 1561 AAAAATCTTCTTACCTTTCAAAATTTTACGATGAGAAACCAGATGTTTATTTTAT 1620
Qy 1621 CTGTGCGGCGTGTGCTGCTCCCTTCTTTTACATGGAACAACTATTTGAGCAAGCA 1680
Db 1621 CTGTGCGGCGTGTGCTGCTCCCTTCTTTTACATGGAACAACTATTTGAGCAAGCA 1680
Qy 1681 TTTATATGAAATTTTGAATGCAATTAATCTACATTTAACTGGAACAACTATTTATC 1740
Db 1681 TTTATATGAAATTTTGAATGCAATTAATCTACATTTAACTGGAACAACTATTTATC 1740
Qy 1741 CCGTAACAGATTTTATTTATGTTTCAAGGGAAGAAATGAAGGATATCTTAATCTA 1800
Db 1741 CCGTAACAGATTTTATTTATGTTTCAAGGGAAGAAATGAAGGATATCTTAATCTA 1800
Qy 1801 TTAACCTGGGAGATTTTATTTATGATTTTGTAAACAGTAGAAGTCAAGCAAGT 1860
Db 1801 TTAACCTGGGAGATTTTATTTATGATTTTGTAAACAGTAGAAGTCAAGCAAGT 1860
Qy 1861 TTAATGATTTGATTAAGAGATTTATTTGAAAAAAATGCAAGTATTAACGAGAAAT 1920
Db 1861 TTAATGATTTGATTAAGAGATTTATTTGAAAAAAATGCAAGTATTAACGAGAAAT 1920
Qy 1921 ATTTATTTCAACAGATTTCTGACATTAATTCAGAAATTTCCAGATATTAATAATTTCT 1980
Db 1921 ATTTATTTCAACAGATTTCTGACATTAATTCAGAAATTTCCAGATATTAATAATTTCT 1980
Qy 1981 CAGTTCAAGAAATGGAACAATTAATTAACAATTCAGAAATGATTTATTTGCAACGAG 2040
Db 1981 CAGTTCAAGAAATGGAACAATTAATTAACAATTCAGAAATGATTTATTTGCAACGAG 2040
Qy 2041 CCCCCTACTTTTATGATTTCAATTTCAAGGAAAAACAATTTTGTTCCTAGAA 2100
Db 2041 CCCCCTACTTTTATGATTTCAATTTCAAGGAAAAACAATTTTGTTCCTAGAA 2100
Qy 2101 AAAAAAGTATGAGCAATGTAATGATCATCAAGTAAAGTTTGAAGAAATTTTACA 2160
Db 2101 AAAAAAGTATGAGCAATGTAATGATCATCAAGTAAAGTTTGAAGAAATTTTACA 2160
Qy 2161 AGATTAATTAATTTTATTAAGAAATATAGATGTTTGTGAAAAAATTAATTTAAGT 2220
Db 2161 AGATTAATTAATTTTATTAAGAAATATAGATGTTTGTGAAAAAATTAATTTAAGT 2220
Qy 2221 TTCTAAGCAACCTTACTTTACATCAATTAATAATTTTGTGAAAAATTAACAAT 2280
Db 2221 TTCTAAGCAACCTTACTTTACATCAATTAATAATTTTGTGAAAAATTAACAAT 2280
Qy 2281 AGTTAAAAATTTAATGAGATCAAGAAATGAATTAATAAAAAAGTATGATTTAGATA 2340
Db 2281 AGTTAAAAATTTAATGAGATCAAGAAATGAATTAATAAAAAAGTATTTAGATA 2340
Qy 2341 TGGCTTATCATTAATTTTCTCAGATTTTCTGAGAGGGAATACAGATTTATCATCTTCT 2400
Db 2341 TGGCTTATCATTAATTTTCTCAGATTTTCTGAGAGGGAATACAGATTTATCATCTTCT 2400

QY 2401 CTCGAGAAATGCAACACATTAGTTCCTTCGAAATACCTGATATATTTTAAATATC 2460
DB CTTGAGGAAATGACACCATAGTTCCTTGAGATACCTGATATATTTTAAATATC 2460
QY 2461 TCGAGATTTATATGTTGATTTTACAAAAGATGAGCAAAAATATTAAGAAAATAGGATATA 2520
DB TCGAGATTTATATGTTGATTTTACAAAAGATGAGCAAAAATATTAAGAAAATAGGATATA 2520
QY 2521 TGAACGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2580
DB TGAACGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2580
QY 2581 ACTGTTAGATTTTATTAAGATGATTCGAGCTTTGAAATCTATTTTACAAAGATTTT 2640
DB ACTGTTAGATTTTATTAAGATGATTCGAGCTTTGAAATCTATTTTACAAAGATTTT 2640
QY 2641 GTTTATTTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2700
DB GTTTATTTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2700
QY 2701 CTTCCACATGATTTTGGGCAATCTTTTATCAAAATGAAAACGAAACGCTATTTATTT 2760
DB CTTCCACATGATTTTGGGCAATCTTTTATCAAAATGAAAACGAAACGCTATTTATTT 2760
QY 2761 AAGTAAATCTAAATGTCAGATGAACTATTTATACAGCAATTTATAGAAAATATGAAATTT 2820
DB AAGTAAATCTAAATGTCAGATGAACTATTTATACAGCAATTTATAGAAAATATGAAATTT 2820
QY 2821 TCAAAATGATTTATCAAAATGAAATTTAAGATATATTAAGTGAAGAAAATCAACATCT 2880
DB TCAAAATGATTTATCAAAATGAAATTTAAGATATATTAAGTGAAGAAAATCAACATCT 2880
QY 2881 TCTCCATTTGCTTTACAGATGATTTCTATGATGAAATTTGCTAAATGCAAGAAATTTAGT 2940
DB TCTCCATTTGCTTTACAGATGATTTCTATGATGAAATTTGCTAAATGCAAGAAATTTAGT 2940
QY 2941 TTTTATTTGCTGAAGATTTAAGAAATGAAATTTAATCTAAATTTAAGAAATTTATTTACT 3000
DB TTTTATTTGCTGAAGATTTAAGAAATGAAATTTAATCTAAATTTAAGAAATTTATTTACT 3000
QY 3001 AAAAAATTAATAGTGAATTTTGTGAGATGATGATGTTAAATTTAATTAATGATGACC 3060
DB AAAAAATTAATAGTGAATTTTGTGAGATGATGATGTTAAATTTAATTAATGATGACC 3060
QY 3061 CGGAATTTTATTTTAAAGTCTTCTGTTGATTTATTTTCCGAGCAAAAGTATG 3120
DB CGGAATTTTATTTTAAAGTCTTCTGTTGATTTATTTTCCGAGCAAAAGTATG 3120
QY 3121 TATTTTATTAATTTTATGAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3180
DB TATTTTATTAATTTTATGAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3180
QY 3181 TAAATATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3240
DB TAAATATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3240
QY 3241 TAGTCACAGTATGTTTGTGAATTAATTTTGAAGATTTTGAAGATTTTGAAGATTTTGAAG 3300
DB TAGTCACAGTATGTTTGTGAATTAATTTTGAAGATTTTGAAGATTTTGAAGATTTTGAAG 3300
QY 3301 CGATATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAAT 3360
DB CGATATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAAT 3360
QY 3361 ATAAAAATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3420
DB ATAAAAATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3420
QY 3421 TGTATATTTATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3480
DB TGTATATTTATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3480
QY 3481 ACTATCTTATTAACAGGCGTCAAAAACAAGTGTGCTTTATGAACTATCTTACGTTAA 3540

DB 3481 ACTATCTTATTAACAGGCGTCAAAAACAAGTGTGCTTTATGAACTATCTTACGTTAA 3540
QY 3541 ATACCACTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3600
DB ATACCACTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3600
QY 3601 AACAAATTTTATTTTCTGCTGCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3660
DB AACAAATTTTATTTTCTGCTGCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3660
QY 3661 GTAGTTATGCTAGCAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3720
DB GTAGTTATGCTAGCAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3720
QY 3721 TTGCTGGATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3780
DB TTGCTGGATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3780
QY 3781 AATTGCTTTACATGAATTTTGGCTGTTATTAATTTCTAGAGATCAAGATCAAGATCAAG 3840
DB AATTGCTTTACATGAATTTTGGCTGTTATTAATTTCTAGAGATCAAGATCAAGATCAAG 3840
QY 3841 GATTTATTTATTTATCAAGAAATTTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 3900
DB GATTTATTTATTTATCAAGAAATTTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 3900
QY 3901 ATGAATATCCGAATATTTACGTTACGGAACCTTGCTCGGAATCATTTAGGCTATATAT 3960
DB ATGAATATCCGAATATTTACGTTACGGAACCTTGCTCGGAATCATTTAGGCTATATAT 3960
QY 3961 CATTTTATTAATTAATCAAGATTTGCTGTTATTTATTTATTTATTTATTTATTTATTTAT 4020
DB CATTTTATTAATTAATCAAGATTTGCTGTTATTTATTTATTTATTTATTTATTTATTTAT 4020
QY 4021 TTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4080
DB TTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4080
QY 4081 CCAATATTTTCAATATTAAGAAACAATGATCCGATTTATTTATTTATTTATTTATTTATTT 4140
DB CCAATATTTTCAATATTAAGAAACAATGATCCGATTTATTTATTTATTTATTTATTTATTT 4140
QY 4141 CTTCAATAGGATTTTGAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4200
DB CTTCAATAGGATTTTGAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4200
QY 4201 GATTTAATTTCACTTATTTGACCAATTTTATTTATTTATTTATTTATTTATTTATTTATTT 4260
DB GATTTAATTTCACTTATTTGACCAATTTTATTTATTTATTTATTTATTTATTTATTTATTT 4260
QY 4261 AACAGATTAATTAACCAACATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4320
DB AACAGATTAATTAACCAACATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4320
QY 4321 ACTGATGATTTCTGAGAAATTTTCTTAACTAATTAAGAAACAATGGAATTAATTAAT 4380
DB ACTGATGATTTCTGAGAAATTTTCTTAACTAATTAAGAAACAATGGAATTAATTAAT 4380
QY 4381 TACAAGAAATTAATGAGCGGCTAGCAGATGCTGGAATTTTCCGATAGAAACATGCAACA 4440
DB TACAAGAAATTAATGAGCGGCTAGCAGATGCTGGAATTTTCCGATAGAAACATGCAACA 4440
QY 4441 GGTAAATTAATTTGCTTTGCTGATCTGATCAATTAATTAAGATTTGCAATGTTTCAAGAGA 4500
DB GGTAAATTAATTTGCTTTGCTGATCTGATCAATTAATTAAGATTTGCAATGTTTCAAGAGA 4500
QY 4501 ATGCAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4560
DB ATGCAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4560
QY 4561 GACGAAACAGGCTATCAAAAACAAGATTAATTTTATTTATTTATTTATTTATTTATTTATTT 4620

Db 4561 GACGAAAAACGGATATACAAAAAGAAAAATAGTATTTTCATGCTCTTAAACGAGAA 4620
Qy 4621 GAGACTGTAAAAAGATTTTTCAGAGATCTAATATAGAAATPANGTTGGTCAAGCTT 4680
Db 4621 GAGACTGTAAAAAGATTTTTCAGAGATCTAATATAGAAATPANGTTGGTCAAGCTT 4680
Qy 4681 TATTCACGAGATTTATTAAGATATAAAATTCCAATTAAATMTAGAAATTTGGTGAG 4740
Db 4681 TATTCACGAGATTTATTAAGATATAAAATTCCAATTAAATMTAGAAATTTGGTGAG 4740
Qy 4741 GATTGCTTTTAAATTGGAGGCTTGAAACATGTAAACAGCTGTAGTACTATA 4800
Db 4741 GATTGCTTTTAAATTGGAGGCTTGAAACATGTAAACAGCTGTAGTACTATA 4800
Qy 4801 GAATATTAATTAATATATGCAATTCGTAACAGTTCGCTTATTAATCAGAAATCTCTATA 4860
Db 4801 GAATATTAATTAATATATGCAATTCGTAACAGTTCGCTTATTAATCAGAAATCTCTATA 4860
Qy 4861 AATAATATTTGATTTAGTCACAAAGTTGAGAAATTAACCCCTTAAAGTTAAAAAGAGATT 4920
Db 4861 AATAATATTTGATTTAGTCACAAAGTTGAGAAATTAACCCCTTAAAGTTAAAAAGAGATT 4920
Qy 4921 AGTCATTATTTTGAAGCAAAAGTTATTAAGAGAGTTAAATGTTTAAACAAATGTAT 4980
Db 4921 AGTCATTATTTTGAAGCAAAAGTTATTAAGAGAGTTAAATGTTTAAACAAATGTAT 4980
Qy 4981 TCACACAGATTTGTTGATATAGATTCTTGCCAAATTTAGTCTATCGAAAAAGATA 5040
Db 4981 TCACACAGATTTGTTGATATAGATTCTTGCCAAATTTAGTCTATCGAAAAAGATA 5040
Qy 5041 CGTAGATATCCATTATTTAAAGCGAAAAAGATTTTATCAAGAAAGCATTTAGTTACGTTG 5100
Db 5041 CGTAGATATCAATTTATTAAGCGAAAAAGATTTTATCAAGAAAGCATTTAGTTACGTTG 5100
Qy 5101 TATTTGATGAATTTTGCCTAAACTATATGTATATATATTAAGAAATTTCAAAGCAG 5160
Db 5101 TATTTGATGAATTTTGCCTAAACTATATGTATATATATTAAGAAATTTCAAAGCAG 5160
Qy 5161 TAGAGGTAAAAATGATAAAAATTTAGTATTTGTTCCAGTTTAAAGTATGAATAATTT 5220
Db 5161 TAGAGGTAAAAATGATAAAAATTTAGTATTTGTTCCAGTTTAAAGTATGAATAATTT 5220
Qy 5221 TAAGTAGTTGTATGAAGCATTAATTAATCAAAATTTAAAAATATATATATTTATGTA 5280
Db 5221 TAAGTAGTTGTATGAAGCATTAATTAATCAAAATTTAAAAATATATATATTTATGTA 5280
Qy 5281 TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGAAATATGCAAAAAAGATA 5340
Db 5281 TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGAAATATGCAAAAAAGATA 5340
Qy 5341 AAAGAGTAAAAATTTTTCCTAATCATAGTGAAGTATCAAAAGCTAAGAAATCATGGA 5400
Db 5341 AAAGAGTAAAAATTTTTCCTAATCATAGTGAAGTATCAAAAGCTAAGAAATCATGGA 5400
Qy 5401 TAAAGCGAGTACAGCTGAATATATGTTGTTGACTGTAGATGTTGTTGATAGTA 5460
Db 5401 TAAAGCGAGTACAGCTGAATATATGTTGTTGACTGTAGATGTTGTTGATAGTA 5460
Qy 5461 GATTAGTAGAAAAATTAATTTTATATATATTAATAAAAGTAGAAGTATTCGTTGTT 5520
Db 5461 GATTAGTAGAAAAATTAATTTTATATATTAATAAAAGTAGAAGTATTCGTTGTT 5520
Qy 5521 TGTACGCTACTTTTTCAGAAAAATATATTTTGAAGTGAATATCAAAATATGATT 5580
Db 5521 TGTACGCTACTTTTTCAGAAAAATATATTTTGAAGTGAATATCAAAATATGATT 5580
Qy 5581 TTGAAGCAATTAATACCGTGAAGCATGGGAGAAAAAATTTATGAAATTTGTATATA 5640
Db 5581 TTGAAGCAATTAATACCGTGAAGCATGGGAGAAAAAATTTATGAAATTTGTATATA 5640
Qy 5641 ATTAATATTTTCTACCTCTGTTGTAAACTATATATAGAAAAAGATACATCAAGATCTTT 5700
Db 5641 ATTAATATTTTCTACCTCTGTTGTAAACTATATATAGAAAAAGATACATCAAGATCTTT 5700

Qy 5701 TTCAAGAGATCATGTTAGGAGAAATTAATCTTTTAAATCTGCATTTATTAAGATA 5760
Db 5701 TTCAAGAGATCATGTTAGGAGAAATTAATCTTTTAAATCTGCATTTATTAAGATA 5760
Qy 5761 TAGATAGATAGTATTTGACTGAACATCTTATTTTATAGAGAGTATACATA 5820
Db 5761 TAGATAGATAGTATTTGACTGAACATCTTATTTTATAGAGAGTATACATA 5820
Qy 5821 CAGTAAATCTTTTAAAGAGGTGTGTTTGGCAATTTGCAAAATTTGCAAAACAGTGA 5880
Db 5821 CAGTAAATCTTTTAAAGAGGTGTGTTTGGCAATTTGCAAAATTTGCAAAACAGTGA 5880
Qy 5881 TAGATATGTTTAAAGCAATATATGTTGAGATTTTGAAGTATCAATTTGTAAGTACTA 5940
Db 5881 TAGATATGTTTAAAGCAATATATGTTGAGATTTTGAAGTATCAATTTGTAAGTACTA 5940
Qy 5941 TAGCTTGGCAATTTTATTTAGCTACTATGCTTAATGTTTAAATACGAAAAACGTCTATT 6000
Db 5941 TAGCTTGGCAATTTTATTTAGCTACTATGCTTAATGTTTAAATACGAAAAACGTCTATT 6000
Qy 6001 TTGACAAATTTTAAATTTTAAAGATCTTATTAATAAAATATTTATTTTAAAG 6060
Db 6001 TTGACAAATTTTAAATTTTAAAGATCTTATTAATAAAATATTTATTTTAAAG 6060
Qy 6061 TATCTAACAAAAATCTTGTCTAAAAATTTTGTATAGAAATTTTTCGAAACAAAGTTT 6120
Db 6061 TATCTAACAAAAATCTTGTCTAAAAATTTTGTATAGAAATTTTTCGAAACAAAGTTT 6120
Qy 6121 TTTAAAAAATATTAAGTATTAATAGAGATATCATGTACTATTAATAATTTCTA 6180
Db 6121 TTTAAAAAATATTAAGTATTAATAGAGATATCATGTACTATTAATAATTTCTA 6180
Qy 6181 TAAATGTACCTATATATATATAGAAAAATTTATCTAAATGTATATAGATTTGTA 6240
Db 6181 TAAATGTACCTATATATATATAGAAAAATTTATCTAAATGTATATAGATTTGTA 6240
Qy 6241 ATCAGACTAACAAACATATAGAGATTTCTTGTGTGATGACGTTAGTACGATATTCGG 6300
Db 6241 ATCAGACTAACAAACATATAGAGATTTCTTGTGTGATGACGTTAGTACGATATTCGG 6300
Qy 6301 AAGAAATTTGTTTGAAGATATGCAAGAAAGATATGTCATTTCTTATTTTAAAAAGAGA 6360
Db 6301 AAGAAATTTGTTTGAAGATATGCAAGAAAGATATGTCATTTCTTATTTTAAAAAGAGA 6360
Qy 6361 ACGCGGGCTATCAGATGCCCCGTAATTAAGCATTAAGTGCAGAGGGAATCTACTAG 6420
Db 6361 ACGCGGGCTATCAGATGCCCCGTAATTAAGCATTAAGTGCAGAGGGAATCTACTAG 6420
Qy 6421 CTTTATTAAGCTCAGATGATTTTATTCATTCGAGATTCAATCCAAAGTTTACAGAAACAA 6480
Db 6421 CTTTATTAAGCTCAGATGATTTTATTCATTCGAGATTCAATCCAAAGTTTACAGAAACAA 6480
Qy 6481 TTGAAGAGAGAAATGCCCTTGTGCGAGTGTCTGTTATGATAGAGTATGCTTCCGGGG 6540
Db 6481 TTGAAGAGAGAAATGCCCTTGTGCGAGTGTCTGTTATGATAGAGTATGCTTCCGGGG 6540
Qy 6541 ATTTCTTAAAGAGAGCCGCTTCTTCAAAATCAGGCTGTCTGAAGCGGAGAAATGTTT 6600
Db 6541 ATTTCTTAAAGAGAGCCGCTTCTTCAAAATCAGGCTGTCTGAAGCGGAGAAATGTTT 6600
Qy 6601 GTAAAAAGCTGTAGAGGCGGATGCTCATTGCTTGTGTTGCTGTAAATPAACTCTATA 6660
Db 6601 GTAAAAAGCTGTAGAGGCGGATGCTCATTGCTTGTGTTGCTGTAAATPAACTCTATA 6660
Qy 6661 AAAAAGACTATTTGAAGATTTTTCATTTGAAAAAGGTAAGATTCATGAAGATGAATACT 6720
Db 6661 AAAAAGACTATTTGAAGATTTTTCATTTGAAAAAGGTAAGATTCATGAAGATGAATACT 6720
Qy 6721 TCAATTAATGCTGCTCTATAGAGTTAGAAAAAGTGCATATAGTTAAGAGTGTCTACT 6780
Db 6721 TCAATTAATGCTGCTCTATAGAGTTAGAAAAAGTGCATATAGTTAAGAGTGTCTACT 6780

QY 6781 ATTATGTGACCGAAGAAATAGTATCAAACTTCAGCATGACTGACCATCGCTTCATT 6840
| | | | |
Db 6781 ATTATGTGACCGAAGAAATAGTATCAAACTTCAGCATGACTGACCATCGCTTCATT 6840
QY 6841 GCTTACTGGAATTTCAAAATGAGAGAAATGGAATCTCTATGAAAGTAGAGATTAAGAC 6900
| | | | |
Db 6841 GCTTACTGGAATTTCAAAATGAGAGAAATGGAATCTCTATGAAAGTAGAGATTAAGAC 6900
QY 6901 TCTTACTAGAGTGTATCGTTCAATTTTGTAGCCTTGGCTTTGTTTGGCAATATA 6960
| | | | |
Db 6901 TCTTACTAGAGTGTATCGTTCAATTTTGTAGCCTTGGCTTTGTTTGGCAATATA 6960
QY 6961 ATCATGTGTGAGCAACAGCAAAAGAGCTT 6992
| | | | |
Db 6961 ATCATGTGTGAGCAACAGCAAAAGAGCTT 6992

RESULT 2
US-09-767-041-29
; Sequence 29, Application US/09767041
; Patent No. US2002005168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 6992
; TYPE: DNA
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: CPS1
US-09-767-041-29

Query Match 100.0%; Score 6992; DB 3; Length 6992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCAAACGAATGGCATTTATGATATGATAGAGTTGCAATTTCTGCAATCTTAA 60
| | | | |
Db 1 ATGCCAAACGAATGGCATTTATGATATGATAGAGTTGCAATTTCTGCAATCTTAA 60
QY 61 CAAGTCATATACCAATGCTGATTTAAATGCTTCTGAAATTTTATCATATGATGTTTC 120
| | | | |
Db 61 CAAGTCATATACCAATGCTGATTTAAATGCTTCTGAAATTTTATCATATGATGTTTC 120
QY 121 ATTATTTGCAATTTTATATCTGCTATGCGAGTTGAATTTAGATAGAGGTATGTA 180
| | | | |
Db 121 ATTATTTGCAATTTTATATCTGCTATGCGAGTTGAATTTAGATAGAGGTATGTA 180
QY 181 TAGAGTTGAAAAAACAATTAACTATGATATGATATTTGCAATTTTCTTACGGCAGTAT 240
| | | | |
Db 181 TAGAGTTGAAAAAACAATTAACTATGATATGATATTTTCTTACGGCAGTAT 240
QY 241 CATTTTGTGAGAAATATTTGCACTTTCAGACGTGTCGCTGATTTTACATTA 300
| | | | |
Db 241 CATTTTGTGAGAAATATTTGCACTTTCAGACGTGTCGCTGATTTTACATTA 300
QY 301 TAAACTCTGTTTGTGATACCTATTTAAAGTATTTAAAGCAAGTTTAAAGATGCTTTC 360
| | | | |
Db 301 TAAACTCTGTTTGTGATACCTATTTAAAGTATTTAAAGCAAGTTTAAAGATGCTTTC 360
QY 361 TATTTTGAACATCTATCAAAAAAAGAGCATTTCAATTACACGGCTGAACGATGGAAA 420
| | | | |

Db 361 TATTTTGAACATCTATCAAAAAAAGAGCATTTCAATTACACGGCTGAACGATGGAAA 420
| | | | |
QY 421 ATATGCAAGTTTATTTGATATCAATTAACAAATTTCAAAATTTCTTGGCATTTGGTAG 480
| | | | |
Db 421 ATATGCAAGTTTATTTGATATCAATTAACAAATTTCAAAATTTCTTGGCATTTGGTAG 480
QY 481 TTTTAAAGTACGAATAATGATTAATTAATTTATATACCGCTCTATTAATTTCTGTGAG 540
| | | | |
Db 481 TTTTAAAGTACGAATAATGATTAATTAATTTATATACCGCTCTATTAATTTCTGTGAG 540
QY 541 AACCTATAGAGTTTCAACAGGAAAGTGTGACACAGTCTTTATTAATCTACCAAGTG 600
| | | | |
Db 541 AACCTATAGAGTTTCAACAGGAAAGTGTGACACAGTCTTTATTAATCTACCAAGTG 600
QY 601 AGTTTGTAGCGTAAGCAATTTGTTTCAATTTTGAATTTGATGATGATGTAAGC 660
| | | | |
Db 601 AGTTTGTAGCGTAAGCAATTTGTTTCAATTTTGAATTTGATGATGATGTAAGC 660
QY 661 TTGATATTAATTCATTCGTTTCTGCTTGAAGAAACAAATTTCAATCTGCTAGTG 720
| | | | |
Db 661 TTGATATTAATTCATTCGTTTCTGCTTGAAGAAACAAATTTCAATCTGCTAGTG 720
QY 721 ACCATGACATTTGATCTTTTTCACAAATTTTATAAGCTTATCATATCATGATGAAC 780
| | | | |
Db 721 ACCATGACATTTGATCTTTTTCACAAATTTTATAAGCTTATCATATCATGATGAAC 780
QY 781 GACTTTGATATATCTGAGCGGTAGTGGGTATTTATTTGATGATGTTCTATTT 840
| | | | |
Db 781 GACTTTGATATATCTGAGCGGTAGTGGGTATTTATTTGATGATGTTCTATTT 840
QY 841 TGTATGTTCAATTTATCTGAGAGATGATGAGACCGGTATTTTGTCTCAGAAAAGAGTTG 900
| | | | |
Db 841 TGTATGTTCAATTTATCTGAGAGATGATGAGACCGGTATTTTGTCTCAGAAAAGAGTTG 900
QY 901 GACAGATGACCGCATATTTATCTATCTACAGTTTGCATGATGATGATGATGATGAG 960
| | | | |
Db 901 GACAGATGACCGCATATTTATCTATCTACAGTTTGCATGATGATGATGATGATGAG 960
QY 961 AGCGCAAAAAAGCTTGTCTGAGCAAAACCAATGCAAGGTGATGTTTAAATG 1020
| | | | |
Db 961 AGCGCAAAAAAGCTTGTCTGAGCAAAACCAATGCAAGGTGATGTTTAAATG 1020
QY 1021 GAAAAACGATCTAGATTAATCTCAATTTGAGCAATTTGATGATGATGATGATGATGAG 1080
| | | | |
Db 1021 GAAAAACGATCTAGATTAATCTCAATTTGAGCAATTTGATGATGATGATGATGATGAG 1080
QY 1081 AGTTACACAGTTTATTAATGTTTAAATTTGCGATATGATGATGATGATGATGATGAG 1140
| | | | |
Db 1081 AGTTACACAGTTTATTAATGTTTAAATTTGCGATATGATGATGATGATGATGATGAG 1140
QY 1141 CTACAGTTGATGATTTGAAAAATATATCTCTGCTCAAAAGAGCAATGATTTTAAAC 1200
| | | | |
Db 1141 CTACAGTTGATGATTTGAAAAATATATCTCTGCTCAAAAGAGCAATGATTTTAAAC 1200
QY 1201 CAGGATTAACAGTCTCTGAGAGTTAGTGTGATGATGATGATGATGATGATGATGAG 1260
| | | | |
Db 1201 CAGGATTAACAGTCTCTGAGAGTTAGTGTGATGATGATGATGATGATGATGATGAG 1260
QY 1261 TAGTTGCTTGAATTTAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
| | | | |
Db 1261 TAGTTGCTTGAATTTAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 TATTAAGACAGTGAAGTTGATTTGTTGAGAGAGAGATTAATGATGATGATGATGATGAG 1380
| | | | |
Db 1321 TATTAAGACAGTGAAGTTGATTTGTTGAGAGAGAGATTAATGATGATGATGATGATGAG 1380
QY 1381 TTTGTTGTGCTGCTCTCAAGGGGCAATTTGATGATGATGATGATGATGATGATGATGATGAT 1440
| | | | |
Db 1381 TTTGTTGTGCTGCTCTCAAGGGGCAATTTGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 GGAAGAGAGAGAGAGCTTTTGGGTAACTTTGATTAAGAGATGCAAGAGTCTTTTGA 1500
| | | | |

Db 1441 GGAAGAAAGAACGTTTTTGGGTAACTTTGATTAAGAGATGACAGAGTCCTTTTGA 1500
Qy 1501 AGAATGAAAAATGTAATCCATGTTACTTCCAAAGATGCAATCTCAATTAATTTAGTA 1560
Db 1501 AAAAAAGAAAAATGATCCATGTTACTTCCAAAGATGCAATCTCAATTAATTTAGTA 1560
Qy 1561 AAAATCTTCTTAGCTTCAAAATTTTACGTGATAGAAACCAAGTGTATTTATTCAT 1620
Db 1561 AAAATCTTCTTAGCTTCAAAATTTTACGTGATAGAAACCAAGTGTATTTATTCAT 1620
Qy 1621 CTGTGCGCGCGCTGCTGCTCCCTTTTACATCGGAAAACTATTTGGAGCAAAACGA 1680
Db 1621 CTGTGCGCGCGCTGCTGCTCCCTTTTACATCGGAAAACTATTTGGAGCAAAACGA 1680
Qy 1681 TTTATTTTGAAGATTTGATTCGAGTTAATTAATCTCAATTAATGAAAACTAGTTATTC 1740
Db 1681 TTTATTTTGAAGATTTGATTCGAGTTAATTAATCTCAATTAATGAAAACTAGTTATTC 1740
Qy 1741 CCGTAAACAGATATTTTATTTGTTCACTGGGAGAGAAATGAAAGATATTCCTMAATCTA 1800
Db 1741 CCGTAAACAGATATTTTATTTGTTCACTGGGAGAGAAATGAAAGATATTCCTMAATCTA 1800
Qy 1801 TTAATCTGGGAGATTTTATTTTAAATGATTTTGTAAACATGACATGACAAACAGTT 1860
Db 1801 TTAATCTGGGAGATTTTATTTTAAATGATTTTGTAAACATGACATGACAAACAGTT 1860
Qy 1861 TTAATCTGGGAGATTTTATTTTAAATGATTTTGTAAACATGACATGACAAACAGTT 1920
Db 1861 TTAATCTGGGAGATTTTATTTTAAATGATTTTGTAAACATGACATGACAAACAGTT 1920
Qy 1921 ATTTATTTCAAAACAGATATTTCTGATATTTCCAGATATTTGCAAGTATTAATAATTTCT 1980
Db 1921 ATTTATTTCAAAACAGATATTTCTGATATTTCCAGATATTTGCAAGTATTAATAATTTCT 1980
Qy 1981 CAGTTAACAAGAAATGAGAACATATATTAACAATCAAGATGTTTCCACGAGG 2040
Db 1981 CAGTTAACAAGAAATGAGAACATATATTAACAATCAAGATGTTTCCACGAGG 2040
Qy 2041 CCCCCGTAATTTTATGAAATTCATTTCCAAAGAAAAACAATATTTGTTCTTACACA 2100
Db 2041 CCCCCGTAATTTTATGAAATTCATTTCCAAAGAAAAACAATATTTGTTCTTACACA 2100
Qy 2101 AAAAAAGATGAGAACATGATTAATGATCAAGATGAGTTGTGAAGAAATTTTACA 2160
Db 2101 AAAAAAGATGAGAACATGATTAATGATCAAGATGAGTTGTGAAGAAATTTTACA 2160
Qy 2161 AGATTAATTAATTTTATTTATTAAGAAAAATATGATGATTTGTTGAAAAATTTATGAAGT 2220
Db 2161 AGATTAATTAATTTTATTTATTAAGAAAAATATGATGATTTGTTGAAAAATTTATGAAGT 2220
Qy 2221 TTTCTAAGCAACATTAATTTTACATCAATTAATTAATTTTGTGAAGAAATTAACAAT 2280
Db 2221 TTTCTAAGCAACATTAATTTTACATCAATTAATTAATTTTGTGAAGAAATTAACAAT 2280
Qy 2281 AGTTGAAAAATTTATGAGATCAAGAAAAATGATTAATTAATTAATTAATTTGATTA 2340
Db 2281 AGTTGAAAAATTTATGAGATCAAGAAAAATGATTAATTAATTAATTAATTTGATTA 2340
Qy 2341 TGGCTTATCAATTTTCTCAGATTTTACTGAGAGGATACAGATATTTATCTCTCT 2400
Db 2341 TGGCTTATCAATTTTCTCAGATTTTACTGAGAGGATACAGATATTTATCTCTCT 2400
Qy 2401 CTGAGAGAAATGACACCATTAAGTTCCCTTCAAGATTAATTAATTAATTTTAAATTC 2460
Db 2401 CTGAGAGAAATGACACCATTAAGTTCCCTTCAAGATTAATTAATTAATTTTAAATTC 2460
Qy 2461 TCAGATTTTATGTTGAATTTTACAAAAAGATGACAAAAATTAATTAATTAATTTAGATA 2520
Db 2461 TCAGATTTTATGTTGAATTTTACAAAAAGATGACAAAAATTAATTAATTTAGATA 2520
Qy 2521 TGAACGATTAATGTTACAGATTAATTTCTTAATATATCAGAAAAAACTATATGTAATGT 2580
Db 2521 TGAACGATTAATGTTACAGATTAATTTCTTAATATATCAGAAAAAACTATATGTAATGT 2580

Qy 2581 ACTGTTAGAAATTTTATTAAGATGATGACCTTTTGAAATATCTAATTTTCAAAAGATGTT 2640
Db 2581 ACTGTTAGAAATTTTATTAAGATGATGACCTTTTGAAATATCTAATTTTCAAAAGATGTT 2640
Qy 2641 GTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700
Db 2641 GTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700
Qy 2701 CTTCACATGATTTTGTGCAATTTCTTATCAATTAATTAATTAATTAATTAATTAAT 2760
Db 2701 CTTCACATGATTTTGTGCAATTTCTTATCAATTAATTAATTAATTAATTAATTAAT 2760
Qy 2761 AAGTAATCTAAATGTCAGATGACATTTATTAAGACAAATTAATTAATTAATTAAT 2820
Db 2761 AAGTAATCTAAATGTCAGATGACATTTATTAAGACAAATTAATTAATTAATTAAT 2820
Qy 2821 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
Db 2821 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
Qy 2881 TCTCTAATGTCCTTTACAGATGATTTCTATGATGAAATTCCTAAATGCAAGAAATTTAGT 2940
Db 2881 TCTCTAATGTCCTTTACAGATGATTTCTATGATGAAATTCCTAAATGCAAGAAATTTAGT 2940
Qy 2941 TTTTATTTGCTAGAAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
Db 2941 TTTTATTTGCTAGAAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
Qy 3001 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3060
Db 3001 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3060
Qy 3061 CCGAATATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3120
Db 3061 CCGAATATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3120
Qy 3121 TATTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3180
Db 3121 TATTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3180
Qy 3181 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3240
Db 3181 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3240
Qy 3241 TAGTCAAGATGATTTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3300
Db 3241 TAGTCAAGATGATTTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3300
Qy 3301 CCAATTAATTTGATTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3360
Db 3301 CCAATTAATTTGATTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3360
Qy 3361 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3420
Db 3361 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3420
Qy 3421 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3480
Db 3421 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3480
Qy 3481 ACTATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
Db 3481 ACTATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
Qy 3541 ATACCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3600
Db 3541 ATACCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3600
Qy 3601 AACAATTTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3660
Db 3601 AACAATTTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3660

OY	3661	5FAGTTTAAACGTGCACATATTTAAATTAATGCTGTATAGGAGATATATAGCTGAAAAT	3720
Db	3661	GTAGTTTAAACGTGCACATATTTAAATTAATGCTGTATAGGAGATATATAGCTGAAAAT	3720
OY	3721	TTGCTGTGATAAAAAGCTAAATAGTAATTTGGTAATCTACTATTAATTTTAATATCTG	3780
Db	3721	TTGCTGTGATAAAAAGCTAAATAGTAATTTGGTAATCTACTATTAATTTTAATATCTG	3780
OY	3781	AATTCCTTTACATGAAAAATTTTGCTGTTTAAATCTAGAGAAATCAAGTAACGAGCTA	3840
Db	3781	AATTCCTTTACATGAAAAATTTTGCTGTTTAAATCTAGAGAAATCAAGTAACGAGCTA	3840
OY	3841	GATTATATATTTATCTACAGGAAGTATGTGTAAGTAATTAAGTAACAAATTTTATTTGGAT	3900
Db	3841	GATTATATATTTATCTACAGGAAGTATGTGTAAGTAATTAAGTAACAAATTTTATTTGGAT	3900
OY	3901	ATGGAATATCCGAATATTCAGTTACGGGAACTTGCTGCGAAGCATTCAGGCTATATAT	3960
Db	3901	ATGGAATATCCGAATATTCAGTTACGGGAACTTGCTGCGAAGCATTCAGGCTATATAT	3960
OY	3961	CATTTTATTAATAATCAGAAATAGTTGGGTTGATTTTACATGATGTTTTCTTTTATATG	4020
Db	3961	CATTTTATTAATAATCAGAAATAGTTGGGTTGATTTTACATGATGTTTTCTTTTATATG	4020
OY	4021	TTATTAATAAAAAAGTTATGAGTTAATGGGAAAACGACATATTTTATTTTATCATCATATAG	4080
Db	4021	TTATTAATAAAAAAGTTATGAGTTAATGGGAAAACGACATATTTTATTTTATCATCATATAG	4080
OY	4081	CCATATTTTCAATATATGAAACAATATAGATCCGATTAATATATATTTAGTACTATTCCTTT	4140
Db	4081	CCATATTTTCAATATATGAAACAATATAGATCCGATTAATATATATTTAGTACTATTCCTTT	4140
OY	4141	CTTCATATGATTTTGGATTAATATATATTTTAAAAAGATATGAGACAAAAATGAAAT	4200
Db	4141	CTTCATATGATTTTGGATTAATATATATTTTAAAAAGATATGAGACAAAAATGAAAT	4200
OY	4201	GATTTAATTTCAATTATTTGTAACAATTTATATATGTCMAAGATATCTTGATTAATGTATT	4260
Db	4201	GATTTAATTTCAATTATTTGTAACAATTTATATATGTCMAAGATATCTTGATTAATGTATT	4260
OY	4261	AAACGATATTTAATCCAAACATATATCTAATTTAGAGTTATTTCTGTAATATGATGAAAT	4320
Db	4261	AAACGATATTTAATCCAAACATATATCTAATTTAGAGTTATTTCTGTAATATGATGAAAT	4320
OY	4321	ACTGATGATTCGAGAAAAATTTGCTTAACATATATATGAAAGAAAGATGGAATTAATAT	4380
Db	4321	ACTGATGATTCGAGAAAAATTTGCTTAACATATATATGAAAGAAAGATGGAATTAATAT	4380
OY	4381	TACAAGAAAAATTAATGGCGGTCTAGACAGATCTCGAAATTTCCGACTTAGAACATGCACA	4440
Db	4381	TACAAGAAAAATTAATGGCGGTCTAGACAGATCTCGAAATTTCCGACTTAGAACATGCACA	4440
OY	4441	GCTAATATATATGCTTTTGTGCGATTCTGATGACATATATAGAAATTTGGAATGTTGGAAGA	4500
Db	4441	GCTAATATATATGCTTTTGTGCGATTCTGATGACATATATAGAAATTTGGAATGTTGGAAGA	4500
OY	4501	ATGCAATGATATATTAATCGAGTATATATGCGAATATGCGAAGATATGATTTTGTATAGTA	4560
Db	4501	ATGCAATGATATATTAATCGAGTATATATGCGAATATGCGAAGATATGATTTTGTATAGTA	4560
OY	4561	GACGAAAAACGGGTATACAAAGAAAAAAGAAATATGTAATTTTCAATGCTTATACGAGGAA	4620
Db	4561	GACGAAAAACGGGTATACAAAGAAAAAAGAAATATGTAATTTTCAATGCTTATACGAGGAA	4620
OY	4621	GAGACTGTAAAGAAATTTTGTGCGAGATCTAATATAGAAATTAATGTTTGGTGCACGCTT	4680
Db	4621	GAGACTGTAAAGAAATTTTGTGCGAGATCTAATATAGAAATTAATGTTTGGTGCACGCTT	4680
OY	4681	TATTCACAGATATATTAATAAGATATATAAATTTCCAAATTAATATAGAAATTTGGTGAG	4740
Db	4681	TATTCACAGATATATTAATAAGATATATAAATTTCCAAATTAATATAGAAATTTGGTGAG	4740
OY	4741	GATTTGCTTTTAAATTTGGAGGCTTGTGAACAATGTAAACGATGATGATTTGATTAATCTAGA	4800

Db	4741	GATTTGCTTTAAATTTGGAGGCTCTTGACCACTTAAACGCTGTAGTAGTTGATCTAGCA	4800
Qy	4801	GAATATATATTAATTAATTAATGTCATTGCTAACAGTTGCTTAATTAATCAGAAATTCCTATA	4860
Db	4801	GAATATATATTAATTAATTAATGTCATTGCTAACAGTTGCTTAATTAATCAGAAATTCCTATA	4860
Qy	4861	AATAATATTAATTAATTAATGTCACAAAGTTGGAGAAATTAACCCCTTTAAAGTTAAAGAGATTT	4920
Db	4861	AATAATATTAATTAATTAATGTCACAAAGTTGGAGAAATTAACCCCTTTAAAGTTAAAGAGATTT	4920
Qy	4921	AGCATATATTTGGATGCAAAAGTAAATTAAGAGAGGTTAAATGCTTTAAACCAAAATGAT	4980
Db	4921	AGCATATATTTGGATGCAAAAGTAAATTAAGAGAGGTTAAATGCTTTAAACCAAAATGAT	4980
Qy	4981	TCAAACAGATTTTGGATTAATGAGTCTTCGCCAATAATGAGCTTTATTCGAAAAAGATA	5040
Db	4981	TCAAACAGATTTTGGATTAATGAGTCTTCGCCAATAATGAGCTTTATTCGAAAAAGATA	5040
Qy	5041	CGTAGATATCCATTATTAATAAGCAAAAAGATATTTATCAAGAAAGCAATTAAGTTACGTTG	5100
Db	5041	CGTAGATATCCATTATTAATAAGCAAAAAGATATTTATCAAGAAAGCAATTAAGTTACGTTG	5100
Qy	5101	TATTTGATGAATTTTGCGCTTAAACCTATATGTAATGTAATTAAGAAATTTCAAAACGAG	5160
Db	5101	TATTTGATGAATTTTGCGCTTAAACCTATATGTAATGTAATTAAGAAATTTCAAAACGAG	5160
Qy	5161	TAGAGGTAAATGAGATTAATTAATGCTGTATATGTTCCAGTTTAAATGAGATTAATTT	5220
Db	5161	TAGAGGTAAATGAGATTAATTAATGCTGTATATGTTCCAGTTTAAATGAGATTAATTT	5220
Qy	5221	TAACTAGTTGTATGAAGAACCTATTAATCAAAATTAATAAAATATAGAAATTAATTAATGA	5280
Db	5221	TAACTAGTTGTATGAAGAACCTATTAATCAAAATTAATAAAATATAGAAATTAATTAATGA	5280
Qy	5281	TAGATGATGCTCTGTAGATGATTTCTGCTAAATATATGCAAGGAATATGCAAAAAAGATA	5340
Db	5281	TAGATGATGCTCTGTAGATGATTTCTGCTAAATATATGCAAGGAATATGCAAAAAAGATA	5340
Qy	5341	AAAAGATTAATTTTTCATATATGCTAATATGCTGAGATATCAAAATGCTAGAAATCATGGA	5400
Db	5341	AAAAGATTAATTTTTCATATATGCTAATATGCTGAGATATCAAAATGCTAGAAATCATGGA	5400
Qy	5401	TAAACGCGAGTACAGCTGAAATATATTAATGTTTGTGACTGCTGATGATGTTGTTGATGTA	5460
Db	5401	TAAACGCGAGTACAGCTGAAATATATTAATGTTTGTGACTGCTGATGATGTTGTTGATGTA	5460
Qy	5461	GATTAGTGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5520
Db	5461	GATTAGTGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5520
Qy	5521	TGTACGCTACTTTTCGAGAAATTAATTAATTAATTTTGAAGGAATTAATCCAAATTAATGAT	5580
Db	5521	TGTACGCTACTTTTCGAGAAATTAATTAATTAATTTTGAAGGAATTAATCCAAATTAATGAT	5580
Qy	5581	TTGAAGCAATTAATACCGTGCAGACATGCGAGAAAAAATTTTAATGAAATTTGTAATTA	5640
Db	5581	TTGAAGCAATTAATACCGTGCAGACATGCGAGAAAAAATTTTAATGAAATTTGTAATTA	5640
Qy	5641	ATAATATTTTCTACTCTGCTGTTGTAACCTATTAAGAAAGATATACATACAGATCTTT	5700
Db	5641	ATAATATTTTCTACTCTGCTGTTGTAACCTATTAAGAAAGATATACATACAGATCTTT	5700
Qy	5701	TTCAAGAGAAATTAATGCTGAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	5760
Db	5701	TTCAAGAGAAATTAATGCTGAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	5760
Qy	5761	TAGATAGAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5820
Db	5761	TAGATAGAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5820
Qy	5821	CAGTAATATCTTTAAAGAGGCTGTTTTTGGCAATTTGGAATTTTGCAGAAATCAAGTGA	5880
Db	5821	CAGTAATATCTTTAAAGAGGCTGTTTTTGGCAATTTTGGCAATTTTGCAGAAATCAAGTGA	5880

Db 5821 CAGTAAATCTTTTAAAGAGGTGTTTTCATGATGCAAAATTTGCAAAACAGTGA 5880
QY 5881 TAGTATGTTTAAAGCAATATATGAGGATTTTACGATCAATTTGTTAAAGTACTA 5940
Db 5881 TAGTATGTTTAAAGCAATATATGAGGATTTTACGATCAATTTGTTAAAGTACTA 5940
QY 5941 TAGCTGGCAAGTATTTTATATAGCTTA CTAA GTTTTAAATACGAAAAACGCTATTT 6000
Db 5941 TAGCTGGCAAGTATTTTATATAGCTTA CTAA GTTTTAAATACGAAAAACGCTATTT 6000
QY 6001 TTGACCAATTTTAAATTTTAAAGAACTTTTAAATTTTAAATTTTAAATTTTAAAG 6060
Db 6001 TTGACCAATTTTAAATTTTAAAGAACTTTTAAATTTTAAATTTTAAATTTTAAAG 6060
QY 6061 TATCTAACAAATTTCTTTGCTTAAATTTTGTATTAAGAAATTTGTTGCAACAAATTT 6120
Db 6061 TATCTAACAAATTTCTTTGCTTAAATTTTGTATTAAGAAATTTGTTGCAACAAATTT 6120
QY 6121 TTTAAAAATATTTATGTTTAAATTTTAAAGAAATTTTGTATTAAGAAATTTGTTGCA 6180
Db 6121 TTTAAAAATATTTATGTTTAAATTTTAAAGAAATTTTGTATTAAGAAATTTGTTGCA 6180
QY 6181 TAAATGCTATATATATATATGTAAGAAATTTTAAATTTTAAATTTTAAATTTTAAAG 6240
Db 6181 TAAATGCTATATATATATATGTAAGAAATTTTAAATTTTAAATTTTAAATTTTAAAG 6240
QY 6241 ATCAGACCTTAAACATATATGAAATTTCTTGTGTAATGACGTAATGACGTAATTTCCG 6300
Db 6241 ATCAGACCTTAAACATATATGAAATTTCTTGTGTAATGACGTAATGACGTAATTTCCG 6300
QY 6301 AAGAAATTTGTTTAAAGCAATATGCAAGAAATTTGCAATTTGTTTAAATTTTAAAGAA 6360
Db 6301 AAGAAATTTGTTTAAAGCAATATGCAAGAAATTTGCAATTTGTTTAAATTTTAAAGAA 6360
QY 6361 ACCGCGGCTTATGATGCTGCTTAAATTTTAAATTTTAAATTTTAAATTTTAAAGAA 6420
Db 6361 ACCGCGGCTTATGATGCTGCTTAAATTTTAAATTTTAAATTTTAAATTTTAAAGAA 6420
QY 6421 CTTTAAATGCTATGATGCTTAAATTTTAAATTTTAAATTTTAAATTTTAAAGAA 6480
Db 6421 CTTTAAATGCTATGATGCTTAAATTTTAAATTTTAAATTTTAAATTTTAAAGAA 6480
QY 6481 TTGAGAGAGAAATGCTTGTGCAATTTTAAATTTTAAATTTTAAATTTTAAAGAA 6540
Db 6481 TTGAGAGAGAAATGCTTGTGCAATTTTAAATTTTAAATTTTAAATTTTAAAGAA 6540
QY 6541 ATTTCTTAAAGCAAGAGCTTCTTAAATTTTAAATTTTAAATTTTAAAGAA 6600
Db 6541 ATTTCTTAAAGCAAGAGCTTCTTAAATTTTAAATTTTAAATTTTAAAGAA 6600
QY 6601 GTTAAAGCTTAAAGAGCTTCTTAAATTTTAAATTTTAAATTTTAAAGAA 6660
Db 6601 GTTAAAGCTTAAAGAGCTTCTTAAATTTTAAATTTTAAATTTTAAAGAA 6660
QY 6661 AAAAAGAACTTAAAGAGCTTCTTAAATTTTAAATTTTAAATTTTAAAGAA 6720
Db 6661 AAAAAGAACTTAAAGAGCTTCTTAAATTTTAAATTTTAAATTTTAAAGAA 6720
QY 6721 TCACTTAACTGCTTCTTAAAGAGCTTCTTAAATTTTAAATTTTAAAGAA 6780
Db 6721 TCACTTAACTGCTTCTTAAAGAGCTTCTTAAATTTTAAATTTTAAAGAA 6780
QY 6781 ATTATGTTGACCGGAAATTTGATCACTTAAAGAGCTTCTTAAATTTTAAAGAA 6840
Db 6781 ATTATGTTGACCGGAAATTTGATCACTTAAAGAGCTTCTTAAATTTTAAAGAA 6840
QY 6841 GCTTAACTGCTTCTTAAAGAGCTTCTTAAATTTTAAATTTTAAAGAA 6900
Db 6841 GCTTAACTGCTTCTTAAAGAGCTTCTTAAATTTTAAATTTTAAAGAA 6900
QY 6901 TCTTAACTGCTTCTTAAAGAGCTTCTTAAATTTTAAATTTTAAAGAA 6960
Db 6901 TCTTAACTGCTTCTTAAAGAGCTTCTTAAATTTTAAATTTTAAAGAA 6960

QY 6961 ATCATGTTGAGCAAAAGCAAAAGAGCTT 6992
Db 6961 ATCATGTTGAGCAAAAGCAAAAGAGCTT 6992

RESULT 3
US-09-900-038A-3
; Sequence 3, Application US/09900038A
; Patent No. US20020142425A1
; GENERAL INFORMATION:
; APPLICANT: Matsubara, Masaki
; APPLICANT: Matsubara, Masaki
; APPLICANT: Matsubara, Masaki
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 766.53
; CURRENT APPLICATION NUMBER: US/09/900,038A
; PRIOR FILING DATE: 2001-09-21
; PRIOR FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6865
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae type 1b
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (617)..(1789)
; NAME/KEY: CDS
; LOCATION: (1816)..(2262)
; NAME/KEY: CDS
; LOCATION: (2265)..(2744)
; NAME/KEY: CDS
; LOCATION: (2843)..(3979)
; NAME/KEY: CDS
; LOCATION: (3982)..(4953)
; NAME/KEY: CDS
; LOCATION: (5009)..(5947)
US-09-900-038A-3

Query Match 10.4%; Score 727.8; DB 3; Length 6865;
Best Local Similarity 59.2%; Pred. No. 1.3e-92;
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;

QY 16 TGGCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75
Db 438 TAGCATGATTTCAACCGCTTGTGTTTATTTTCTGCAAGTTGATGATGATGATGATGATGATGAT 497
QY 76 ATGCTGATTTAAATCG--TTCTGAAATTTTATCATTAATGATGCTTCAATTTTGCAT 132
Db 498 CTCCTCACTTTAAACCAATTAAGATTTATGTTGTTCTATGATGATGATGATGATGATGATGAT 557
QY 133 TTTTATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
Db 558 TTTTATCTTCTGATTTTAAACAGACTTTTGAAGTGTGCTGATGATGATGATGATGATGATGATGAT 617
QY 193 AAACATTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
Db 618 TGGTATGAAATACAGCTTTTACTATATTTTCAATCAAGTTCAATTTTATTTTCTA 677
QY 253 AGAATATTTTCCGACTTTCAAGAGCTGCTGCTGATTTTCACTTAAATTTTCACTTAAATTTTCACTT 312
Db 678 AAAACCTTTTAAACAGAGAGCTTCTTCTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 737
QY 313 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
Db 738 TATTTATATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 797
QY 373 TCTATCAAAAGAGAGATTTCTAATTTCAACGCTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 432
Db 798 CACGAGATCAACAGAGTTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 857

QY 433 TATTGATGACATTAACAAATTCMAAAATCTTGTGATGCTGATGTTAGTGA CAG 492
DB 858 GGAATTAAT---ACGACCATTAATTAATGCTGCTGATCTTGATTCCTGGAAG 914
QY 493 AAATAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 552
DB 915 ATGTGATGATTTGAAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 974
QY 553 TTTCAACAAAGGAGTGTGACACGCTTAAATTAATTAATTAATTAATTAATTA 612
DB 975 AGTTAACTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1034
QY 613 TAAAGCAATTC-----GTTTCAGATTTGAGTTGATTAATTAATTAATTAAT 663
DB 1035 AATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1094
QY 664 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 723
DB 1095 ATGTAGAGGACCTTATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1154
QY 724 ATAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 783
DB 1155 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1214
QY 784 TTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 843
DB 1215 TTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1274
QY 844 TATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 903
DB 1275 TATGTCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1334
QY 904 AGATGAGGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 963
DB 1335 GTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1394
QY 964 GCAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1023
DB 1395 TTTAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1449
QY 1024 AAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1083
DB 1450 CGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1508
QY 1084 TACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1143
DB 1509 TGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1568
QY 1144 CAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1203
DB 1569 CAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1268
QY 1204 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1263
DB 1629 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1688
QY 1264 TTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1323
DB 1689 TAAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1748
QY 1324 TAAAGCAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1372
DB 1749 TCTTAACGCTAAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1808
QY 1373 -----TATGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1426
DB 1809 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1868
QY 1427 GTTAAACCGTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1486
DB 1869 TTTGAAGCCATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1928
QY 1487 AAGAAGCTTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1546

DB 1929 TAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1988
QY 1547 CATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1606
DB 1989 CAAAACCTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2048
QY 1607 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1666
DB 2049 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2108
QY 1667 TGAAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1726
DB 2109 TGTCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2168
QY 1727 AAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1786
DB 2169 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2228
QY 1787 ATATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1846
DB 2229 TTTATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2288
QY 1847 CATGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1906
DB 2289 CATGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2348
QY 1907 ATTAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1966
DB 2349 ATTAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2408
QY 1967 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2026
DB 2409 TGTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2468
QY 2027 ATTTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2086
DB 2469 ATCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2528
QY 2087 TTTGTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2146
DB 2529 GTGTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2588
QY 2147 AGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2206
DB 2589 AAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2268
QY 2207 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2263
DB 2649 ATTATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2708
QY 2264 GAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2323
DB 2709 TCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2768
QY 2324 AGATGCA 2330
DB 2769 AATGAA 2775

RESULT 4
US-10-865-873-3
; Sequence 3, Application US/10865873
; Publication No. US20050064559A1
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsuhide
; APPLICANT: Matsumoto, Masaki
; APPLICANT: Tjijme, Shuji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 766.53
; CURRENT APPLICATION NUMBER: US/10/865,873
; CURRENT FILING DATE: 2004-06-14
; PRIOR APPLICATION NUMBER: US/09/900,038

Db 2109 TGGCTGTAAGACCGTTATATAGAGTTTCGACAGATAGATAACCACTTGACAGG 2168
Qy 1727 AAAACGTATTATCCCTTAACAGATATTTTATTTGTTCACTGGGGAAGAAATGAAGAGT 1786
Db 2169 AAAATTAAGTATCTCTGTAACAGTAATTAATTAATTAATTAATTAATTAATTAATTA 2228
Qy 1787 ATATCTTAATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1846
Db 2229 TTAATCTTAAGGCAATTAATTAAGAGAAATTTTATTAATTAATTTTGTCAAGTAGGACA 2288
Qy 1847 CATGAACAAGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1906
Db 2289 CATGAACAAGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2348
Qy 1907 ATTAACCAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1966
Db 2349 ATTAATCAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2408
Qy 1967 TATTAATAATTTCTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2026
Db 2409 TGGTCAAAATTTCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2468
Qy 2027 ATTTGCCAGGAGGCGGCTACTTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 2086
Db 2469 ATCAACAGGCGGCTCAAGCAAGTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 2528
Qy 2087 TTGTTCTTAAGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2146
Db 2529 GTGTTCTTAAGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2588
Qy 2147 AGAAGATTTTACAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2206
Db 2589 AAAAGATTATTTCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2648
Qy 2207 AAATTAATTAATTTCTTAAGCAAACT--TAACTTATCAATTAATTAATTAATTTTGT 2263
Db 2649 ATATTAAGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2708
Qy 2264 GAAAGATTAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2323
Db 2709 TCTCTTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2768
Qy 2324 AGATGCA 2330
Db 2769 AATTGAA 2775

RESULT 5
US-09-870-759-83
Sequence 83, Application US/09870759
Patent No. US2002017751A1
GENERAL INFORMATION:
APPLICANT: TERMAN, David S
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
FILE REFERENCE: 870759
CURRENT APPLICATION NUMBER: US/09/870,759
PRIOR FILING DATE: 2002-01-14
PRIOR APPLICATION NUMBER: US 60/208,128
NUMBER OF SEQ ID NOS: 166
SOFTWARE: PatentIn version 3.1
SEQ ID NO 83
LENGTH: 17276
TYPE: DNA
ORGANISM: Streptococcus agalactiae
FEATURE:
NAME/KEY: CDS
LOCATION: (7062)..(8207)
OTHER INFORMATION:
US-09-870-759-83

Query Match 10.4%; Score 726.8; DB 3; Length 17276;

Best Local Similarity 59.2%; Pred. No. 2,4e-92;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;
Qy 16 TGGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 75
Db 4765 TACGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4824
Qy 76 ATGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 132
Db 4825 CTCCCAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4884
Qy 133 TTTTATATCTCGATGCGAGTTGAATTTGATTAATTAATTAATTAATTAATTAATTAATTA 192
Db 4885 TTTATCTTCTGATTTTACAGAGACTTTGAGTCTGCTATCTTGAAGAGTTTAA 4944
Qy 193 AAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 252
Db 4945 TGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5004
Qy 253 AGAATTAATTTGCGACTTTCAAGAGTGTGCGGCTGATTTTCACTTAATTAATTAATTAAT 312
Db 5005 AAACCTTTTACAGAGCAAGCTTCTTTTATCTTTATTTATTTGATTAATTAATTAAT 5064
Qy 313 TGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 372
Db 5065 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5124
Qy 373 TCTATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 432
Db 5125 CACGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5184
Qy 433 TATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 492
Db 5185 GGAATTAAT--ACGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5241
Qy 493 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 552
Db 5242 ATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5301
Qy 553 TTTCAACAAGGAGTGTGACCAAGCTTTTAAATTAATTAATTAATTAATTAATTAATTAAT 612
Db 5302 AGTTAACGTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5361
Qy 613 TAAAGCA-----TTGCTTCAATTTTGAATTTGATTAATTAATTAATTAATTAATTAAT 663
Db 5362 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5421
Qy 664 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 723
Db 5422 ATGTAAGGACACTTACCTTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5481
Qy 724 ATNGCATTTGACTTTTCCAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 783
Db 5482 ATAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5541
Qy 784 TTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 843
Db 5542 TTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8601
Qy 844 TAGTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 903
Db 5602 TAGTTCCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5661
Qy 904 AGAATGACGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 963
Db 5662 GTAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5721
Qy 964 GCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1023
Db 5722 TTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5776
Qy 1024 AAACGATCTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1083

Db 5777 AGATGATCCTAGAAATTAATAAATAGAAATTTATTCG-AAAAACAAGCATAGTAGT 5835
Qy 1084 TACCACAGTTTATATGTTTTTATTTGGGATATAGTCTAGTTGGTACAGTCCACCTA 1143
Db 5836 TCCCTCAATTCATATATGTTTTTAAAGGCGATATGTTTACAGAGAACACCCCTCCCA 5895
Qy 1144 CAGTGTAGTAAATTTGAAAAATATACCTCTGTCTCAAAAGAGACGATTTGATTTAAACCG 1203
Db 5896 CAGTGTAGTAAATTTGAAAAATATTTCAACGACAGAGCGCCCTAGTTTAAAGCCAG 5955
Qy 1204 GGATTAACAGGCTCTCGGACAGGTTAGTGTCTGATATATTCACACACTTCGACAGCTAG 1263
Db 5956 GAATCACTGGTTTGGCAAAATATCTGTAGAAATATATCTGATTTTGTATGTAATG 6015
Qy 1264 TTCCGTGTGACTAGATACATTTGATTAATTTGACTATCTGTCAGATTTTAAATTTTAT 1323
Db 6016 TAAAGTTAGATGTTCATATATTCAAAGAAATGCTATTTGGTACGATTTAAAGATTATTC 6075
Qy 1324 TAAAGCAGTGAAGTTGATTTGTTGAGAGAGGAGTAAGTAAAGTA----- 1372
Db 6076 TCCCTAACACTTAAAGTAGTTTACTCGGACAGAGACTAGTAAAGTTAGGTTGAAG 6135
Qy 1373 -----TATGAAAGTTTGTGTTGTCGTTCTTCAAGGAGACATTTGACTCATTTGATTT 1426
Db 6136 GAATATATGAAATTTGTTCTGGTTGTTCAAGTGTGTCATCTAGCACACTTGAACCT 6195
Qy 1427 GTTAAACCGTTTGGAGAGAAAGAACGTTTGGGTAACTTTGATTAAGAGATGC 1486
Db 6196 TTTGAAACCCATTTGGGAAAAAGAAAGATGTTTGGTAACTTTGATTAAGAAATGC 6255
Qy 1487 AAGAGCTTTTGAAGAAATGAAAAAAATGATTCATGTTACTTTCACAACAATGCAATCT 1546
Db 6256 TAGAGATATTTCTAAGAGAGAGATGTATATCATTTGTTCTTTCACAAACCGTATATG 6315
Qy 1547 CATTAATTTAGTAAAAATATCTTTCTTACCTTCAAAATTTTACGTGATGAGAAACGACA 1606
Db 6316 CAAAACTTGTGTAATAAATATCTATTCATCTTTTAAAGTCTTGAAGAAAGAACGACA 6375
Qy 1607 TGTATATTTATTCATCTGTGCGCCGCTGTCGCCCTCTTTTACATGCGGAAACGATTT 1666
Db 6376 TGTATATCATATCATCTGTGTCGTCGTCGTAAGACATCTTTTATATTTGTAAGTTATTT 6435
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGATTTGATTCAGATTAATTAATCTACATTTAATCTG 1726
Db 6436 TGGTTGTAAACCGTTTATATATAGAGTTTCGACAGATAGATTAACCACTTTGACAG 6495
Qy 1727 AAAACTAGTTTATCCGCTTAAAGATATTTTATTTTGTTCAGTGGAGAAAAATGAAGAGT 1786
Db 6496 AAATTTAGTATCTCTGTAAAGATTAATTTATTTGTTCAAGTGGAGAAAAATGAAGAGT 6555
Qy 1787 ATATCTTAATCTATTAACCTTGGGAGATTTTTTTTATGATTTTGTGAAGATAGTACT 1846
Db 6556 TTTATCTTAAGCAATTAATTTAGAGAGAAATTTTATGATTTTGTGACAGTGGGACACA 6615
Qy 1847 CATGAACAACGATTTAATTCATTTGAAGAGATTTATTTGAATTTTGAATTTGAAGT 1906
Db 6616 CATGAACAACGATTTCAACCGCTTTATTAAGAAATTTGATTAATTAAGGACAGGAGTCT 6675
Qy 1907 ATTAACGACGAAATATTTATTTCAACAGATATTTCTGATATATTTGCAAGATTTGCAAG 1966
Db 6676 ATTTGATCAAGAGTTCATTTCAACAGGTTTACTCAAGCTTCGAACCTCAGAAATTTGACAG 6735
Qy 1967 TATTAATAAATTTCTCAGTTACAAAGAAATGGAACATATTAACAATCAAGAGTGT 2026
Db 6736 TGTCAAAATTTCTCTCATATGATGATATGAACTCTTAATGAAGAAAGCTGAAGATGTT 6795
Qy 2027 ATTTGCAACGAGGCGCGCTTATTTATGAAATTCATTTCAAAAGGAAAAAAACAATTA 2086
Db 6796 ATCAACATGCGCGCGCACGCTTTATATGTCAGTATTTCTTTTAAAGGAAATTTACCACTT 6855
Qy 2087 TTTGTTCTTACAAAAAAGATATGTCAGATGATTAATGATCAATCAAGTAGAGTTGTA 2146
Db 6856 GTTGTCTTACGAGAAAGCAGTTTGTGTAACATATCAATGATCAATCAATCAATTTTAA 6915

Qy 2147 AGAAGATTTTACAGATATATATTTTATTTATAGAAAAATATGATGATTTGTGAA 2206
Db 6916 AAAAAAATTTGCCCACTGTATCTCTTGCTGTGATTTGAAGATGATGAAGACTTGGGAA 6975
Qy 2207 AAAATATTTGAAGTTTCTTACAGCAACCTAACCTTACATCAATATATATTTTGTGAA 2266
Db 6976 GCGTTGAAAGAGATATATGCTTACAGAAAAATATTCAGGAAATATATGTTTTGTCTAT 7035
Qy 2267 AGATTAACCAATAGTTGAAAAAT 2292
Db 7036 AAATTAAGAAAAATATATAGTGAAAT 7061

RESULT 6
US-09-751-708A-83
; Sequence 83, Application US/09751708A
; Publication No. US20030157113A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 751708
; CURRENT APPLICATION NUMBER: US/09/751,708A
; PRIOR FILING DATE: 2002-10-15
; PRIOR FILING DATE: 1999-12-28
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 83
; LENGTH: 17276
; TYPE: DNA
; ORGANISM: *Streptococcus agalactiae*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7062)..(8207)
; OTHER INFORMATION:
; US-09-751-708A-83

Query Match 10.4%; Score 726.8; DB 3; Length 17276;
Best Local Similarity 59.2%; Pred. No. 2,4e-92;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy 16 TGGCATTTATTTGATATGATAGACAGTTCGCAATTTTCGCAATCTTAAAGTCATATCCAA 75
Db 4765 TAGCGATGATTCAAAGAGTTGATTTATTTTCTGCAAGTTGACATTAACATTAATTA 4824
Qy 76 ATGCTGATTTAATGCG---TTCGAAATTTTATGCAATTAATGATGCTCATTTTTCAT 132
Db 4825 CTCCTCAATTTTAAAGCAATTAAGATTTATTTGTTGTTCTATTTGATATATATTTGTT 4884
Qy 133 TTTTATATCTCGTATGCGAGTTGAATTTGAGTATAGAGTAATCTGATAGCTTGAAA 192
Db 4885 TTTATCTTCTGTATTTTACAGAGACTTTTGAGAGTGTGGCTATTTGAAGAGTTTAAAA 4944
Qy 193 AAAATTTTAACTATAGTATATATTTTGCATTTTCTTACGCGAGATATCATTTTGTGG 252
Db 4945 TGGTATGAAATACACTTTTACATATATTTTCAATATCAAGTTCAATATTTTATTTTAA 5004
Qy 253 AGAATATTTGCGACTTTCAGACGTCGTCGCGTATTTTCAATTAATTAATCTGCTTT 312
Db 5005 AAAACTCTTTTACAGACGACGACTTCTTTTATTTTACATTTATTTGCTATGAAATTCGATTT 5064
Qy 313 TGGTATACATTTTAAAGTATTTTAAAGCAGTTTAAAGATAGCTTTCTATTTTTCAGAA 372
Db 5065 TATTTATATCTATTTGAATATTTTAAATATTTATGAAATATTTCTTACGCTAAAGTTT 5124
Qy 373 TCTATCAAAAAAAGACGATTTCTATTTTCAACGCGCTGACAGATGGAATAATATGCAATTT 432
Db 5125 CAGAGATACCAAGTGTGTTTGTATTAAGATTAAGATTTCTTATCTCAAAATGACCTTTA 5184
Qy 433 TATTTGATCATCAATTAACAAATTTCAAAAAATCTGTGTGATTTGTTGTTTGGTTCAG 492
Db 5185 GGAATTAAT--ACGACCATTAATTAATATGCTGTCTGTATCTTGAAGTCTCTCTGAAAGG 5241


```

? PRIOR APPLICATION NUMBER: US 60/415,400
?
? PRIOR FILING DATE: 2002-10-02
? PRIOR APPLICATION NUMBER: US 60/438,686
? PRIOR FILING DATE: 2003-01-09
? NUMBER OF SEQ ID NOS: 224
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 79
?
? LENGTH: 17276
?
? TYPE: DNA
? ORGANISM: Streptococcus agalactiae
?
? FEATURES:
? NAME/KEY: CDS
? LOCATION: (7062)..(8207)
?
? US-10-428-817A-79

```

Query Match	10.4%;	Score 726.8;	DB 8;	Length 17276;
Best Local Similarity	59.2%;	Pred. No. 2.4e-92;		
Matches 1366; Conservative	0;	Mismatches 902;	Indels 38;	Gaps 67

OY	16	TGGCATTA	TTGGATATGAT	TAGACAGTTGCAATTTCTGCATCTTTAAACAAGTCATATACC	75
Db	4765	TAGGCATATTTCA	AAACAGTTGATTTATTTTCTGCAAGTTTACATTTAAATTA	4824	
OY	76	ATGCGATTTAAATG	---TTGCGAATTTTATCATATATAGATGTCATATTTTGCAT	132	
Db	4825	CTCCCAATTTTAA	AGCATATAAGATTTATTTGTTTGTCTATATGATACATTTATTTGTTT	4884	
OY	133	TTTTTATATCTG	TATGCGAGTTGATTTGAGTAACTGATATCTGATAGAGTTTAAA	192	
Db	4885	TTTATCTTCTGAT	TTTTTACAGAGCTTTTGGAGTCGTCATCTTGAAAGTTTAAA	4944	
OY	193	AAACATTTA	CTACTATAGTATTAATATTTTGCATTTTCTTACGGCAGTATCATTTTGTGTGG	252	
Db	4945	TGCTATTA	GAATACAGCTTTTACTATTTTTCATATCAAGTTGATATTTTATTTTATTTTA	5004	
OY	253	AGATTAATTTG	GCACCTTTTCAAGACCTGCGTGCCTGATTTTGCATTAATTAACCTCGTT	312	
Db	5005	AAACCTCTTTT	ACAACGACACACACTCTCTTTTACTTTTATTTTGTCTATAGAAATTCGATTT	5064	
OY	313	TGCTATACCTAT	TTTAAAGTAAATTTTAAAGCAGTTTAAAGATAGCTTCTATTTTGCACA	372	
Db	5065	TATTTATCTAT	TGAATTCATTTTAAATATNTATCGAAAATTTCTTACCGTAAGTTT	5124	
OY	373	TCTATCA	AAAAAAGACGATTTCTAATTAACAACGGCTGAACGATGGGAAAATATGCAAGTT	432	
Db	5125	CACGAGAT	CAAGATGTTTGTGATTAACGAATTAAGATTTCTTATCAAAAAATGACCTTTA	5184	
OY	433	TATTTGAT	CACATTAACAATTTCAAAAAATCTTTGCAATTTGGTACTTTTAGGTACAG	492	
Db	5185	GGAAATTAAT	---ACGACCTAATTTATATCGCTGTGTGATCTCTGACCTCTGAAAAGG	5241	
OY	493	AAATGATTA	AAATTAATTTATTTATCATTAACGCTCTATTTATTTCTGTGGAAGACGTATAGAGT	552	
Db	5242	ATTTGTTATG	ATTTGAAACATTAACCTGTTAAGATATATAACAACAAATGCTCTTACTTCAG	5301	
OY	553	TTTCAACA	CGGAGTGTGACACACGCTTTTATAATCTACCAAGTGAATTTTAAAGC	612	
Db	5302	AATTAAC	CGTCTTAACGTTTGAATCAAGCTTTTATTAACAATCCACTTGAATTAATTTGGTA	5361	
OY	613	TAAAGCA	-----TTGCTTTACAGATTTTGAGTTGTAGGTATGATGTAAAGCTTTG	663	
Db	5362	AATACCA	AAATPACAGATATTAATTAAGAATTTGAAGCAATGGAGATGTGCAATGTTA	5421	
OY	664	AATTTAAT	TCATTCGGTTTACGCGCTTGAAGAAAACAAAATAATCCAACTGTAGGTGAC	723	
Db	5422	ATGTAGAG	CACTTACCTTTGATTAATATATAGAGAAAAGCAATCCAACTTTTGAAGAT	5481	
OY	724	ATAGCA	ATTGTAACTTTTCCACAAATTTTATPAGCTTATGATATCATGATGAACGAC	783	
Db	5482	ATAGGT	TTATACATATTTCTATGAATTTCTATTAATATATGTACACTTATAGCAAAACGAT	5541	
OY	784	TTTTGGAT	TAATCTCGAGCGGTAGTGGGTTAATTTATTTGTGTATATGTTTCTATTTTGT	843	

Db	5542	TTTGTGATATCAGGGTGCTATTATATAGTTTGCTCATATGAGCATTTGGCAATTTTTC	5601
Qy	844	TAGTTCAAATTAATTCGTAGAGATGTGGACCGGCTATATTTTGCTCGAAGACGATTGGAC	903
Db	5602	TAGTTCCGCAATATCAGAAAAAGATGGTGGACCGGCTATCTTTTCTCAAAATPAGATAGTTC	5661
Qy	904	AGAAATGACCATATTTTACATTTCTACAGTTTCGATTCGATGTTGATGCTGAGGAC	963
Db	5662	GTAATGGTAGAATTTTGTAGATTCATTAATTCAGATCAATGCGAGTAGATGCAAGAACAA	5721
Qy	964	GCAAAAAAGCTTGTCAGCCAAACAGATGCGAGGGGTATATTTTAAAAATGGCAA	1023
Db	5722	TTAAGAAAGATTATTTAGTTTCACAATCAATGACAGG-----CTAATGTTAAAGTTACA	5776
Qy	1024	AAACGATCTAGATTACTCCAAATTGGACATTTTCATACGCAAAAACAAGTTTACAGACT	1083
Db	5777	AGATGATCTAGATTACTAAAAATAGGAAATTTTATTCG-AAAAACAACATATGATGAGT	5835
Qy	1084	TACCAAGTTTATATATGTTTAAATTTGGCGATATATGATCTAGTTGGTACACGTCACCTTA	1143
Db	5836	TGCTCTCAATCTAATAAGTTTAAAGGCGATATGATTTAGGAGAACACGCTTCCTCCA	5895
Qy	1144	CAGTTGATGAATTTGAAAAATATACCTCGTCAAAAAGACGATTTAGTTTAAACGAG	1203
Db	5896	CAGTTGATGAATATGAAAAAGATATATTCACAGCAAGAGACGCTTACTGTTTAAACGAG	5955
Qy	1204	GGATTACAGGCTCTGCGACAGTTAGTGTGATATATACAGACTTCGACGACGTAG	1263
Db	5956	GAATCACGTTGTTGGCAAAATCTGTGAAATATATATTAATCTGATTTTGTATGAATATG	6015
Qy	1264	TTGCGTTGACTTACATACATTGATATATTTGACCTATCTGTCGATATTTAAATTTTAT	1323
Db	6016	TAAAGTTAGATGTTCAATATATCAATTAAGTGTCTATTTGGTCAATATTTAAGATTATTC	6075
Qy	1324	TAAAGACGTAAAGTTGTATTTGTTAGACAGGAGAAATAGTAAAAAGTA-----	1372
Db	6076	TCCTAACCTTAAGSTAGTTTACTCGGACAGAGCTAATGTAAGSTAGTTTGAAG	6135
Qy	1373	-----TATGAAAGTTTGTGTGTCGGTCTTCAAGGGGACATTTGACTCATCTTTATTT	1426
Db	6136	GAATTAATGAAAAATTTGTCTGTTGGTTCGTAAGGGGTGCTCATCTAGCACACTTGAACCT	6195
Qy	1427	GTTAAACCCGTTTGGAGAGAAAGACGTTTGGGTAAACATTGATPAAAGAGATGC	1486
Db	6196	TTTGAACCCATTTGGGAAAAAGAAAGATGGTTTGGGTAAACCTTGTATPAAAGAGATGC	6255
Qy	1487	AAGAAGCTTTTGAAGATGAAAAAAGTATCCATGTTACTTTCACAACAAATGCAATCT	1546
Db	6256	TAGAGATTTCTTAAGAGAAAGATTTGATATCATTCGTTCTTTCACAAACAAACGTAATGT	6315
Qy	1547	CATTAAATTAAGTAAAAATACCTTTCTTAGCTTCAAAATTTTACGTATAGAGAAACCA	1606
Db	6316	CAAAAACCTTGGTAAAAAATACTATTTAGCTTTTAAAGTCTTGAAGAAACCAAGACCA	6375
Qy	1607	TGTTATATTTTCAATCGTGGCGCGCTGCTGCCCCCTGTTTACATCGGAAAAACATTT	1666
Db	6376	TGTTATATATCATCTGCGTGGCGCTGAGACATCACTCTTTATATTTGGTATGTTATTT	6435
Qy	1667	TGAGACAAAGACGATTTATTTATTTGAAGATTTGATCGAGTTAATTAATCTACATTTAC	1726
Db	6436	TGGTTGTAAGACCGTTTATATAGAGCTTTTTCGACAGATATGATTAACCAACTTTGACAGG	6495
Qy	1727	AAAACTAGTTTATCCCGTACAGATTTTATTTATTTGTTCACTGGGAGAAATGAAGAGGT	1786
Db	6496	AAAATTAAGTATCTGTATACAGATTAATTTATTTGTTCACTGGGAGAAATGAAGAAAGT	6555
Qy	1787	AATTCCTAATCTATTTAACTTGGGAGATTTTATTTATGATTTTGTATACATAGAGAACT	1846
Db	6556	TATATCTTAAGGCAATTTAATTTAGAGGAATTTTAAAGATTTTGTCACTGGGAGACA	6615
Qy	1847	CATGAAACAACAGTTTATTCGATTTGATAAAGAGATTTGATTTATGAAAAAATATGAGAT	1906
Db	6616	CATGAAACGCAATTTACCGTCTTATTTAAAGATTTGATTAAGTTAATAAAGGACAGGTGCT	6675


```

QY      1324 TAAAGACGTGAAAGTGTATGTTGAGAGAGAGTAAGTAAAAA----- 1372
DB      6076 TCCTAACTTAAGAGTATTTACTGGGACAGAGCTAAAGTAAGTTGAAG 6135
QY      1373 -----TATGAAAGTTTGTGTCGTTCTTCAAGGGGACATTGACTCATTGTATTT 1426
DB      6136 GAATATAAGAAAAATTTGTCGTGGTTCAAGTGTGTCTATGACACACTTGAACT 6195
QY      1427 GTTAAACCGTTTGGAGAGAAAGAGTGTGTTGGGTAACTTGTATTAAGAGATGC 1486
DB      6196 TTTGAAACCCATTTGGGAAAAAGAAATAGCTTTGGGTAACTTTGATTAAGAGATGC 6255
QY      1487 AAGAGCTCTTTGAGAGATGAAAAATGTATCCAGTCTTCTTCAACAATCGCAATCT 1546
DB      6256 TAGAGATATCTTAAGAGAGAGATTTGTATATCATCTTCTTCAACAACCGTATATGT 6315
QY      1547 CATTAATTTAGTAAAAATCTTTCTTACGTTTCAAAATTTTACGTATGAGAAACCA 1606
DB      6316 CAAAAACTGTAAAAAATATCTATCTAGCTTTTAAGTCTTCAAGAAAAAGAACCA 6375
QY      1607 TGTATATTTTCACTGTCGCGCGCGTGTGCTGCTCCCTCTTTACATCGGAAAACTAT 1666
DB      6376 TGTATATCAATATCTGTCGCGCGCTGTGAGATCACTCTTTATATTTGTATAGTTAT 6435
QY      1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGATTAATTAATCTACATTAAC 1726
DB      6436 TGTGTTGTAAGCCGTTTATATAGGTTTTCGACAGGATTAACCAACTTTGACAG 6495
QY      1727 AAAACTAGTTATCCGTTAACAGATATTTTATTTGTTCACTGGGAAAGAAATGAAGGT 1786
DB      6496 AAAATAGTGTATCTGTAAACAGATAAATTTATGTTCACTGGGAAAGAAATGA 6555
QY      1787 ATATCTTAATCTAATTAAGTGGGAGTATTTTATATGATTTTGTAAACATGAGAACT 1846
DB      6556 TTAATCTTAAGCAATTAATTAAGAGAAATTTTATATGATTTTGTCACTGGGAGCA 6615
QY      1847 CATGAACAAGTTTATTCGATTTGATTAAGAGATTTGATTAAGAAAAATGAAGT 1906
DB      6616 CATGAACAAGATTTCAACCGTCTTATTAAGAGTTGATAGATTAAGAAAGGACAGGT 6675
QY      1907 ATAACGAGCAAAATTTTATTTCAACAGATTTCTGACTATATTTCCAGATATTTGCA 1966
DB      6676 ATGATCAAGAAAGTTTCAATTCACAAAGGTTTACTCAAGCTTCGAACTCGAAAT 6735
QY      1967 TATAAAAAATTTTCTCACTTCAAAAGAAATGGAACAATATATTAACAATGAGAGT 2026
DB      6736 TGTCTAAATTTCTCTCATATGATGATATGAATCTTATCAATGAAGAGCTGAGATT 6795
QY      2027 ATTGCGACGAGGCGCGCTACTTTATGATTTCAATTCACAAAGGAAAAACAATTA 2086
DB      6796 ATCAACATGCGCGCGCCAGGAGCTTTATGTCAGTTATTTCTTAAAGGAAATTAACG 6855
QY      2087 TTGTTTCTTGAACAAAAAAGTATGTGAACATGTAATGATCATCAAGTAGAGTTGTA 2146
DB      6856 GTTGTCTTGAAGAAAGCAAGTTGTGAACATATCATGATCATCAAAATACAAATTTTA 6915
QY      2147 AGAAGATTTTCAAGATATATATTTTATTAAGAAAAATATAGATGATTTGTTGA 2206
DB      6916 AAAAAAATTCGCCCTGTATCCCTTGCTGATTAAGATGATGATGATGATGATGATG 6975
QY      2207 AAAATTTGTAAGTTTCTAAGCAAACTTAATTCATCAATTAATTTTATTTTGTGA 2266
DB      6976 GCGTTGAAAAAGATATATGCTACGAAAAATATACAGGAAATATATGATTTGTTGAT 7035
QY      2267 AGATTAACAATAATGTTGAAAAAT 2292
DB      7036 AAATTAAGAAAAATTAATAGTGAAT 7061

```

RESULT 9
 US-10-192-280-1
 ; Sequence 1, Application US/10192280
 ; Publication No. US2004009574A1

```

; GENERAL INFORMATION:
; APPLICANT: Nanibushan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: STREPTOCOCCUS AGALACTIAE CAPSULAR POLYSACCHARIDE SYNTHESIS
; TITLE OF INVENTION: GENES
; FILE REFERENCE: 475412001300
; CURRENT APPLICATION NUMBER: US/10/192,280
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 25020
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-192-280-1

Query Match      10.3%; Score 721.4; DB 6; Length 25020;
Best Local Similarity 58.9%; Pred.No. 1.6e-91;
Matches 1389; Conservative 0; Mismatches 926; Indels 42; Gaps 7;

QY      16 TGGCATATTTGATATGATAGAGTGGCAATTTCTGCAATCTTAAACAATATATCA 75
DB      5154 TAGCGATATTTCAACAAGTTGTGTTTATTTCTGCAAGTTGACATTAATTAATTA 5213
QY      76 ATGCTGATTTAAATCGTCTGGAATTT---TATCATATATATGCTTCATATTTTGA 131
DB      5214 CTCCTCAATTTTAAAGCAATAAAGATTTATGTTGTTCTAATATGATATATATATGTC 5273
QY      132 TTTTATATATCTCGATATGCAATTTGATAGATATGATATGATAGTTTGA 191
DB      5274 TTTTATCTTCTGATTTTTCACAGACCTTTTGAAGTGTGCTATCTTGAAGTTTAA 5333
QY      192 AAAACTTAATCTAATATGATTAATTTTGAATTTTCTTACGAGATATCAATTTTGTG 251
DB      5334 ATGATATTAATTAATGATTTTACTATATTTTCAATCAAGTTCAATATTTTATTTT 5393
QY      252 GGAATATATTTGCACTTTCAAGAGCTGTCGCTGTATTTCAATTAATTAATCTGCT 311
DB      5394 AAAACTCATTTTACACAGACAGACCTTCTTTTCTTTTATGCTATGATTAATGAT 5453
QY      312 TTGATATCTTCTTATTAAGTATTTTATAGAGTTTAAAGATAGCTTCTATTTTGA 371
DB      5454 TTAATGATCTATGATTAATCAATTTTAAATTTATATGAAATATTTCTTAACGTA 5513
QY      372 ATCTATCAAAAAAGACGATTTCAATTTACAGCGCTGAACGATGGAATAATGCAAGT 431
DB      5514 TACGAGATACCAAGTTGTTTGAATGAAGATTTCTTTATCAAAAAATGACCTTT 5573
QY      432 TTAATTTGAATCAATTAACAAATTTCAAAAAATCTTTGATGATTTTATAGTTA 491
DB      5574 AGGATTAAT---ACGACATATATTTATGCTGCTGATCTGATCTGCTGAAAA 5630
QY      492 GAAATATGATTAATTTATTTATCATTAACGCTCTATTTCTGTGGAAGAGATATAG 551
DB      5631 GATTTGATATGATTTTGAACATTAATCTGTTAAGATATTAACAAAGATGCTTTACT 5690
QY      552 TTTTCAACAAGGAGAGTGTGACACGCTTTATTAATTTCAACAGTATTTTGAAC 611
DB      5691 GAGTTAATCTGCTTAACTGTGATCAAGCTTTATTAATCAATCCATGATTAATTTG 5750
QY      612 GTAAAGCAATTCG-----TTTCAATTTTGAATTTGATGATTTGATGATGAT 662
DB      5751 AAATTAACAATAATGATATTTATTAATGATTAATGATTAATGATTAATGATTAAT 5810
QY      663 GATTTAATTTATCTGCTTTTACTGCTTGAAGAAACAAAAATTTCAACTGTTAGG 722
DB      5811 AATGTAAGGCACTTGTGATTAATATAGGAAAAAGCAATTCAACTTTTGAAG 5870
QY      723 CATAGATTTAATCTTTTCACAAAATTTTATATAGCTGATATGATGATGATGATG 782
DB      5871 TATAGTGTATTTATCATATTTCTATGAAATTTCTATTAATATATGATCACTTATAG 5930

```


558 ACAAGGAAGTGTGACACGCTTTTAATCTAACAGTGAAGTTTGTAGACGTAAAG 617
6318 GCNAATTCACCAAGTGAACAGATATTTGT---TGCAGAAAATCCACAGATATTTAT 6374
618 CAATTGTTTTCAGATTTTGAAGTTTGTAGTATTGATGAAGCTGATATTAATTCATTC 677
6375 GAATTGTGCTATTGCTTAAATTTGTAGGAATTCACACAGATGCTGTTGGGAATTA 6434
678 GGTTTTACGCGTGAAGAAACAAAATCCAACTGCTAGTGACATGACATGATTAAT 737
6435 TCGGACTTCTATGTGGAAATAGTGTCTAAAGAAAAGTGTGATACGCTTCATTAAG 6494
738 TTTTCACAAATTTTATTAAGCTTAAGCTATCATGATGAAGAGCTTTGGATATATCT 797
6495 ACAGCATTCATATTTCTAAATTTCCGTCAGATGCTTTAAAGCTTATGATATGCA 6554
798 GGAGCGGTAGTCGGGTTAATTAATTTGTGTATAGTTTCTAATTTTGTAGTTCCATTA 857
6555 ATAGCTTTAGTTGGCTTAGTATTACTGGTATTTGTAGCATTAATATACACCGATTAAT 6614
858 CGTAGAGATG---TGGACCGCTATTTTGTCTAGAAAAGATGGAATGGAACGC 914
6615 AAGAAACATTCACACAGGACCTTAATCTCAACAAAACGTTGTGTAAGAAACGCTAAA 6674
915 ATATTACATCTCAACAAGTTTGCATGATGATGTGATGCTGAGGAGCGCAAAAAGAC 974
6675 GTTTTGAATTTTCAAAATTTTGAAGCATGTACACGATGCCAAGAACGCAAAAAGAA 6734
975 TTGCTCAGCCAAAACAGATGCAAGGCTGGGTATGTTTAAATGGGAAAACGATCTTA 1034
6735 TTACTAACCAAAATGATTTGTGATCTGACTTAATGTTTAAAGATG---ATGATGACCTTC 6792
1035 GAATTACTGCAATTTGACATTTTATGCAAGAAAACAGTTTATGACAGATTAACAGCTT 1094
6793 GTATCTTCCATTTGACATTAAGTTACGTGATTTGTCATT---GATGAATTAACCAATTT 6851
1095 TATTAATGTTTAAATTTGCGATATGAGTCTAGTTGTATGATGCTGACCTACAGTTGATGA 1154
6852 ATTAATGTCTTAAAGGTAAATGTCTGTGTGGGACAGCTCCACCAACGCTTGAAGAA 6911
1155 TTTGAAAAATATATCTCTGTGTCAAAAGACAGATGAATGTTTAAACCGAGGATTAACGT 1214
6912 TATCATCATATAGTTATCATCATTTCAACGATTTGCAACCAACGAGAAATTAACGT 6971
1215 CTGTGCAAGTTAGTGTCTGATTAATTCACAGACTTTCGACGACGATGCTGTTGAC 1274
6972 TTATGCAAGTTAGCGGTCTGATGACATTAACGACTTTGAAGAAAGCTGATGACATTAAT 7031
1275 TTAGCATATCATGATTAATTTGACATATCTGTCAGATATTAATTTTAAAGACAGT 1334
7032 ATGAAGTATATCCAAACTGAGCATATGAGATATTAATTAATTTATGCCAAAACATTT 7091
1335 AAGTGTATTTGTGAGAGAGGAAATGAATTA---AGTATGAAGATTTGTTGTGTC 1392
7092 GAGGTGTATTAAGAGAGGAAATGAATGAATTAATTAAGATTTGTTAGTAG 7151
1393 GTTCTCAGGGGACATTTGATCTCATTTGATTTGTTAAACCGTTTGGAGAGAAAG 1452
7152 GTTCTCTGTGACATTTGACATTTGAATATGCTTAAACCCCTTTTGAAGTGAACATA 7211
1453 AAGTTTGGGTAACTTTGATTAAGAGATGCAAGAGCTTTGAAGATGAAGAAA 1512
7212 GCGGTTCGAGGTATCATTTGATTAAGAGACGCAAGAGGTGTAAAGATGAAGAAAAT 7271
1513 TGTATCATGTATCTTTCAACAATGCAATCTCATTAATTTAGTAAATATATCTTCT 1572
7272 TTTATCGGTATATTTCCGATCAAGAAATTTTAAGATTTGTTAAAGAACCTTCT 7331
1573 TAGCTTCAAAATTTTACGTATGAGAAACCAATGTTATTTATTTATCTGTGTGCGCG 1632
7332 TAGACATTTGAATTTTAAAGAAAAGAAAACCTGACGTATATTTATCTAGAGACGCG 7391

1633 TTGCTGTCCCTCTCTTTTACATCGGAAAACATATTTGGAGCAAGACATTTATATTAAG 1692
7392 TAGACATTTCAATTTTATCTGAGTAACGTTTGGAGCAAGACGTTATATATGAG 7451
1693 TATTTGATCGATTAATTAATCTCATTTACTGCAAAAATGATTTATCCGTAAACATA 1752
7452 TATTTGATAGAAATAGATTAACCGACTGATCGGAAAGTTGGTTATTCAGTGAACATA 7511
1753 TTTTATTTGTTCAGTGGAGAAATGAAGAGATATCTTAATCTTATTAATCTTGGGA 1812
7512 AATTTATGTTCAAGGGAGAGATGAAGAACTGTATCTCAAGCTTATTAATCGGGA 7571
1813 GTATTTTAAATGATTTTGTAAACAGTAGAACTGATGAACAACAGTTAAATCGATGAT 1872
7572 GTATTTTAAATGATTTTGTAAACAGTAGAACTGATGAACAACAGCTTATTAATCGGGA 7631
1873 AAAAGAGATGATTTATTTGAAGAAAATGAAGATTAACCGACGAATATTTATCAAC 1932
7632 TAGGAAGTTGATCGTTTAAAGAAAAGATTTATTAACAGATGAGTTTATTAACAG 7691
1933 AAGATATTTCTGACTATATTCAGAAATTTGCAATTAATTAATTTCTCATTAACAA 1992
7692 AGTTTTCATCTATGAGCTCAATACGTGACGAAATATTAATTTCTTATTCGA 7751
1993 AATGCAATATATTAACAATCAGAGTATTTTCCACGAGGCGCGTACTTT 2052
7752 AATGCAAGATTAATGATATCTGACATTAATTAATTAACAGATGCTGACGACAT 7811
2053 TATGAATCATATTCAAAGGAAAACAAATATTTGTTCTTACAGCAAAAAGATATG 2112
7812 CATGGAGCATTTGTTAAAGGAAAACCAATTTGTTCCAAACAGGAAAGTTGG 7871
2113 TGAACATGTAATGATCATCAAGTATGAGTTGTGAAGAAATTT---TACAAATTAATA 2169
7872 AAGGATGTAATGATCATCAAGTATGAGTTGTGAAGAAATTT---TACAAATTAATA 2169
2170 TATTTTATTTATGAAGAAATATGATGATTTGTTGAAAATTTATTTGAAGTTTCAAGA 2229
7932 TATGTTGTGATGAAGAAATTAATGAATTTGAAAATTTATTTATTTATTTATTTGT 7991
2230 AACTAATCTTACATCAATTAATTAATTTTGTGAAGATTTAAACAAATGTTGAATA 2289
7992 AAGTGAAGTTCAATTCGACACCACTAAGATTTATTAATGCTCAATTAACAAAG---AAA 8047
2290 ATTTATGAGATCAAGAAATGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2349
8048 TAGAAAGTTGTGATGATGATGATCTTAAGAAAGATTAATTAATTTGTTGTTGAGAA 8107
2350 ATTAATTTTCTGATTTTATCTGAGAGGATTAATTAATTAATTAATTAATTAATTAAT 2409
8108 ATCTCTCTCTGACAGTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8167
2410 ATGCAACACATTAATCTTCTGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2468
8168 ATGAATTAATCGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8227
2469 TATATGTAATTTTACAAAGATGACCAAAAATTAATTAATTAATTAATTAATTAATTAAT 2528
8228 AAGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8287
2529 TTAATGTTACAGATTAATTTCTTAATA-----TATCAAGAAAATTAATTAATTAAT 2580
8288 ATTAATGAGGCGGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8347
2581 ACTGTTTAATTAATTTTAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2640
8348 TAACTTCTGAAACCTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8407
2641 GTTTATGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700
8408 GTTTTGTGAGAA--AAAGACATCTTTTATTAAGAAATTAATTAATTAATTAATTAATTAAT 8466
2701 CTTCACATGATTTTGTGCAATTTCTTTATCAATGAAGAAACGAAACAGCTTATTTATTT 2760

Db 8467 GTTCTTAACTTAAGTACGAGAAACATGTGATATCAGCAAAATTTATATTA 8526
Qy 2761 AAGTAATCTAAATCCAGATGAACTTTATACAGCAATATAGAAAATATGAATTT 2820
Db 8527 TCAAGGGGTTATATAGTAAATTCATTCAAAATTTAGTGTCTCATTTATCCA 8586
Qy 2821 TCAATAGATTAATCTAAATATGAAATTTAGATATATTAAGTGAATAATCAACATC 2880
Db 8587 ACGAATTTTGTGTCCGTTTAAATGCAACACAGAAATGGAAATTACTAAAATATCT 8646
Qy 2881 TCTCTAATGCTTTACAGATGATTCATGTGATGAAATGCTAAATGCAAGAAATTTAGT 2940
Db 8647 TATTCATTCATCATTAATGATTCACCTTGATG-----GTAATGCGTGT 8691
Qy 2941 TTTTATTTCTGAGAAAGTTAAATATGAAATATCAATTAATTTAAAGAAATTTACT 3000
Db 8692 AGTGCATATTAATTAAGAAATTTATTCATTAAGATTAATCTCGTATCTTATTCAT 8751
Qy 3001 AAAAAATTAATAGTGAATTTGAGAGATTAATGATGTTAAATTTAAATATGACC 3060
Db 8752 AAATATTTAGGTGAGGCTCTTATGCTA-AAATCAAGCTATTAATTAAGAAATGATATTT 8810
Qy 3061 CGGAATATTTTAAATTTTAACTTCTGTTGATTTATTTTATTCAGAGCAAAAGATG 3120
Db 8811 TTCAAGAGAGATTTTGTTAATAATTTGATTTTGTAT-----ATATATATAT 8864
Qy 3121 TATTTTATTAATTTTATGAAATTAATTTTATTCATTAATTAATTTTGAATTAAC 3180
Db 8865 TCTGTACTCTAATTTAAAGATGGAGCTGAAGCAGTAAATTAATATATGTTGAGC 8924
Qy 3181 TAAATTAATTAATGAAATTTTATGTTTATTAATGCTATATATATGTTTGTTCAG 3240
Db 8925 ATTTTATTAATGATCTAATGCGATTTGTTAAATTTCAAGAAATTTTCATTTAA 8984
Qy 3241 TAGTCAAGATGATGTTTGAATTAATTTTGAAGATTTTGCAGATTTTACTGCTC 3300
Db 8985 TGAATCTTAAATTAATGCTT-----TAAATTTGATTAATGAGCTATTAATTTATGTTG 9040
Qy 3301 CCATTAATTTGATTTATGCAATATATGATTAATTAATTTGATTCATTTAAATTTGAT 3360
Db 9041 GCAAAAGTACACGTTTGTGTTTAACTGCAATGCAATGAGCTTAAATATGTTATG 9100
Qy 3361 ATTAATTAATTAATTAATGATCTTTTATGTTTATGTTTATTAATGATATCTGAT 3420
Db 9101 AAATGCAATTAATTAATTAATGCTTTTGAATGATTAATCT-----CT 9144
Qy 3421 TGTATATTTCAAAATGGAAGAATTTGATTTTGAAGACAGACCTTATAGACATG 3480
Db 9145 TTTCTATTAATGAGTCTTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 9204
Qy 3481 ACTATCTTATTAAGAGGCTGCAAAACAGATTTGCTGCTTATGATGATCTTAACT 3540
Db 9205 TATAGGATGCGCAGTTTATGAGGAAGCATATTTGTTATGGA--CATCCGAATCAG 9262
Qy 3541 ATACCACTAATTAATTAATGATCCGTTATCTTTGCACTTATTAATTAATTAATTAAT 3600
Db 9263 CGAGAGTCTTAAACATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9322
Qy 3601 AACAAATTTTCTGCTGCTGCTTATTAATGATGATTTTAAAGGATGCGAATG 3660
Db 9323 ATATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9382
Qy 3661 GTAGTTATGCTAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3720
Db 9383 CGGCTACAGGTTCTTGATCGAGTATTAATTAATTAATTAATTAATTAATTAATTAAT 9430
Qy 3721 TTGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3780
Db 9431 TTCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9490
Qy 3781 AATGCTTACATGAATTTGCTGTTATTAATTAATTAATTAATTAATTAATTAATTAAT 3840

Db 9491 GGGCTGTTTATGACGCTATTTATAGGTAATTTT-----ACATTAAGATTCACAAAT 9544
Qy 3841 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3900
Db 9545 TAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9604
Qy 3901 ATGAATATCCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3960
Db 9605 ATTAATATCCGCACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9664
Qy 3961 CATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4020
Db 9665 GTTTATTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9724
Qy 4021 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4080
Db 9725 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9784
Qy 4081 CCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4140
Db 9785 GCACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9842
Qy 4141 CTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4200
Db 9843 CTCTTAATTTCAATGTTAACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9898
Qy 4201 GATTAATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
Db 9899 CATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9958
Qy 4261 AACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4320
Db 9959 GAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10018
Qy 4321 ACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4380
Db 10019 AATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10074
Qy 4381 TACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4440
Db 10075 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10099
Qy 4441 GGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4500
Db 10106 GGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10150
Qy 4501 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4560
Db 10151 TGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10210
Qy 4561 GACGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4620
Db 10211 AAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10270
Qy 4621 GAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4680
Db 10271 ACTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10330
Qy 4681 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4740
Db 10331 TTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10390
Qy 4741 GATTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4800
Db 10391 TTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10450
Qy 4801 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4860
Db 10451 CTATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10504
Qy 4861 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4920
Db 10505 TATCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10564

QY 4921 AGTCATTATTTGANGCAAAAGTTATTAAGAGAGGTTAAATGTTTAAACAAATGTAT 4980
 DB 10565 ACTTATTTATTTGTTTAAAGAGAAAATAATCTCGCAAAATTTTACAGGTTTTAAC 10624
 QY 4981 TCAACAGATGTTTGGATAATGAGTCTTGCCAAATATTAAGTCTTATGCAAAAAGAAATA 5040
 DB 10625 CGAATTTGTAATGTTTAAAGTGTCTTGAAATGAACAAGCATTATTCGTAGCAACATCA 10684
 QY 5041 CGTAGATTCATTATTAAGCGAAAAGATATTATCAAGAAAGATTTAGTTAGTGG 5100
 DB 10685 AAGAAATGAATAATTCGCCAAGAGCTATCTGTGGGAAAGTATGCTAAACCATGGCA 10744
 QY 5101 TATTTGATGAATTTTCGCTTAACTATATGTAATGTTATATAGAATTTCAAAACGAG 5160
 DB 10745 TATTTATTTTATTT-----AAAAAGTATTAATAGAAATAGATATCTATGTAGGAG 10796
 QY 5161 TAGAGGTAAATATGATTAATTAATTAAGTATGTTTCCAGTTTAAATGATGAATAAT 5220
 DB 10797 ATATGTGGATATGATTAATCAGTATTTATGTTCTGTATACAAATGTTGAAATAATACA 10856
 QY 5221 TAAAGTGTGTATAGAAAGCATTTATTAATCAAAATTTTAAATATTAAGAAATTTATTA 5280
 DB 10857 TTGCTAAGTGTGAGCTCTTGTAGTTAACCAACATATTTAAATATAGAAATCTCTTAA 10916
 QY 5281 TAGATGATGCTCTGTAGATGATTCGTCTAAATATGCAAGAAATATGCAAAAAAGATA 5340
 DB 10917 TTGATGATGATCTTACAGCAAAATCATATTCGATATGTAAGAAATATGCTGAGTTGAT 10976
 QY 5341 AAGAGTAAATAATTTTCTACTAATCATAGTGAAGTATCAAAATGCTAGAAATCATGAGAA 5400
 DB 10977 CTGGAATTAAGCTTTTCTTAAAGAAATGCGCGCTTTCTAGCGCTCGAAATCTAGGTC 11036
 QY 5401 TAAAGCGAGTACAGCTGAATATATGTTTGTGCTCTGATGATGTTGTGATAGTA 5460
 DB 11037 TTCTACATGTTCAAGAGAGTACGTTGTGTTGATGATCAGATGACTTGTATCACCAA 11096
 QY 5461 GATTAGTAAATAATTTATTTTAAATTTTAAATTAAGAAAGTATTTATCTGTTGT 5520
 DB 11097 AATATGTGAACTTTTATATCAACTTACTTAAATCTTAACTGAGATTTAGCT----- 11149
 QY 5521 TGACGCTACTTTTTCAGAAATATTAATTAATTTTGAAGTGAATATATCAATATATGAT 5580
 DB 11150 --TCTGTAAGTCTGTTATTAACATTTTGAATTAAGAGTGTATTAAGATATCGAATTTATCT 11207
 QY 5581 TTGAAGCAATTTAATCCGTGACAGACATGCGAAGAAAAATTTTATGATTTGATATTA 5640
 DB 11208 TTATCAAAATPACATCAGATGAAAGCTTAAGAAAAATCTTTTGAAGGAGTAAATT 11267
 QY 5641 AATATATTTTCTACTCTGTTGTAAACTATATTAAGAAAAAGATATACATTAACAGATCTT 5700
 DB 11268 GTTATCTTTTTCAAAAATATTTTAAATGTAAACTATTAAGAGCTCGATTTGATGAAA 11327
 QY 5701 TTCAAGAAATCAATGTTAGAGAAAGATTTACTTTTAAATCTGCAATTTTAAAGAAATA 5760
 DB 11328 GTTTAGATACGACGAGAGAGCTTTGTTTATTTATTAACAACCTTAAGAACTAATATTTG 11387
 QY 5761 TAGATAGATTTAGTTATTTGACTGAACATCTTTATTTTATAGAGAGATATCTAAGTA 5820
 DB 11388 CATCTATGATGACGACTGTGACAGATTAATTTTATATCTTA--GAGAAAGATCTTTTACA 11446
 QY 5821 CAGTAATTTCTTTTAAAGAGTGTGTTTTCGCAATTTGGAATTTGCAAAAAAGAGTA 5880
 DB 11447 AATTAAGACTGACTCATCAAGATTAATGATGATTCATTAAGTGTGCGAAATTTTATTA 11506
 QY 5881 TAGATTTGTTTAAAGCAATATATATGTTGAGATTTTGAAGTATCAATTTTAAAGATCTA 5940
 DB 11507 AGAATTTGCAACACGCAAAAAAATTTGAATTTGAAGTA--AATTAATGAATATCAT 11564
 QY 5941 TACGTTGGCAAGTATTTTATTAATGCTTACTAATGTTTAAATACGAAAAAGCTATTT 6000
 DB 11565 TAAAGGAGTGTGAGTCTTGAAGTATTCATTAATTAATGAACTTAAGATGAGTTGAAG 11624

QY 6001 TTGACAAATTTTATTTTATAGAAATCTTATTAATAAAATATTAATTTTAACTGTTAAAG 6060
 DB 11625 AATATTAATATATCACTAGAGAG--AGTTAAAGGTTTAAATTTGATCATTAAGTTCAA 11683
 QY 6061 TATCTAACAAATCTTTGTCTTAAAAATTTTGTATTAAGAAATGTTTGAACA----- 6114
 DB 11684 TATCTAACCTTTAAAAATTTTATAGATTAATTTATTAAGTATAGTCTAGATTAAGTT 11743
 QY 6115 AAGTTTTTAAAAAATATTAATGTTTAAATTAAGAAAGATATCATGATTAATTAAGTTAA-- 6172
 DB 11744 ACAATCTTAAAAAATTAATAGGATTCCTGGAAGAGATTCATGATTTAAATAGTAAC 11803
 QY 6173 ----AATTTCTTAATGCTATATATATTAATTAATGTAAGAAATATTAATCTAATGATAG 6228
 DB 11804 CTCTGTTTCAATTTATTTTCCAAATTTATTAATGATGAAAAATTTTGAACAGTCTCTA 11863
 QY 6229 ATAGCATGTAATCAAGACCTTACAACA 6266
 DB 11864 CTGAGTGTATACATTAATTTATGACA 11891

RESULT 11

US-10-804-408-166
 ; Sequence 166, Application US/10804408
 ; Publication No. US20040253617A1
 ; GENERAL INFORMATION:
 ; APPLICANT: FANRONG, KONG
 ; APPLICANT: GILBERT, GWENDOLYN
 ; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
 ; FILE REFERENCE: 675002-2001
 ; CURRENT APPLICATION NUMBER: US/10/804,408
 ; CURRENT FILING DATE: 2004-03-19
 ; PRIOR APPLICATION NUMBER: PCT/AU02/01281
 ; PRIOR FILING DATE: 2002-09-18
 ; PRIOR APPLICATION NUMBER: AU PR 7749
 ; PRIOR FILING DATE: 2001-09-19
 ; NUMBER OF SEQ ID NOS: 162
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 166
 ; LENGTH: 2226
 ; TYPE: DNA
 ; ORGANISM: Streptococcus agalactiae
 US-10-804-408-166

Query Match 9.3%; Score 652.4; DB 8; Length 2226;
 Best Local Similarity 59.6%; Pred. No. 3.5e-82;
 Matches 1225; Conservative 0; Mismatches 791; Indels 38; Gaps 6;

QY 16 TGGCATTAATTTGATATGATATGACGTTGCAATTTCTGCAATCTTAACAAGCATATACCA 75
 DB 182 TAGCGATGATTCAAAAGAGTGTGTATTTTCTGCMAAGTTTGACATTAACATTAATTA 241
 QY 76 ATGCTGATTTAATGCG---TTCTGAAATTTTATCATTAATGATGTTTCAATATTTTGCAT 132
 DB 242 CTCGCAATTTTAAAGCAATTAAGATTTATTTGTTTGTCTATATGATATATTAATTTGCT 301
 QY 133 TTTTATATCTGATGCCAGTTGAATTTGATATAGAGTAATCTGATAGAGTTGAAA 192
 DB 302 TTTATCTTTCTGATTTTACAGAGACTTTTGGAGTGTGCTATCTTGAAGAGTTTAAA 361
 QY 193 AAACATTTAATTAATGATATATTTGCAATTTTCTTACGCGAGATCAATTTTGTGG 252
 DB 362 TGGATTAATTAATGAGTGTATTAATTTTCAATATCAATCAATGATTAATTTTATTTTAA 421
 QY 253 AGAATATTTTGCACCTTCAAGAGTGTGCTGCTGATTTCACTTAATTAATCAATGCTT 312
 DB 422 AAACATTAATTAATGACAGACTTCTTTTCTTTTATTTGCTATGAATTCGATTT 481
 QY 313 TGGATATCTATTTAAAGTATTTATTAAGCAGTTTAAAGATAGCTTTCTATTTTGCACAA 372
 DB 482 TATTAATATCTATGATTTCAATTTTAAATATTAATGAAATATTTCTAGGCTAATGTTT 541
 QY 373 TCTATCAAAAAAAGACGATTTCTAATTAATCAACGCTGAACGATGGAAAAATATGCAAGTT 432

Db	542	CACGAGATACCAAACTGTTTGGATTAACGAATAAGGATTCCTTATCAGAAATGACCTTTA	601
Qy	433	TATTTGAAATCACATTAACCAAAATTCGTTGTCATTTGGTAACTTTAGGTACAG	492
Db	602	AGAAATAAAT---ACGACCATAATTATTCGCTGCTGATCTCTGGACCTCTCGAAAGG	658
Qy	433	AAATGATATAAATTAAATTTATCATCCGCTCTATTTCTGTGGAAGAACGTATAGGT	552
Db	659	ATTGTTATGATTTGAAACATTAACCTGTTAAGGATTAATAACAAATATCTCTTACTTCAG	718
Qy	553	TTTCAACAAGGAGAGTGGTCGACGACGCTTTATTAATCTACCAAGTAGTTTTAG---	609
Db	719	AGTTAACTCGCTTAACCTGTGATCAAGCTTTTATTAACATACCACTTGAATTTATTTGGTA	778
Qy	610	-----ACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTAGTATGATGAAGCGTTG	663
Db	779	AATACCAAAATACAGATATTTATTAATGACATTTGAAGCAATGGAGATGTTGTCAATGTTA	838
Qy	664	ATATTAAATTCATTCGGTTTACTCGCTTGAAAAACAAAAATCCACTGCTAGTGACC	723
Db	839	ATGTAGAGGCACTTGTGCTTGTATTAATAAGGAAAAACGAATCCAAACTTTTGAAAGAT	898
Qy	724	ATAGCATGTAACTTTTCCAAATTTTAAATTAACCCAGTATATCATGATGAAGAACGAC	783
Db	899	ATAGGTTATTAACATATTCATGAAATTCATATAAATATATAGTCACTTATACCAAAACGAT	958
Qy	784	TTTTGGATATATCTCGAGCGGTAGTCGGGTTAATTAATTTGTGTATGTTCTTATTTTGT	843
Db	959	TTTTGGATATATCATGGGTGCTATTAATAGTTTGTCTCATATGTGCACTTGTGGCAATTTTC	1018
Qy	844	TAGTTCCAATTAATTCGTAGAGATGTGTGACCGGCTATTTTGTCTCAGAAACGATGTGAC	903
Db	1019	TAGTTCGCCCAATTCAGAAAGATGTGTGACCGGCTATTTTCTCAAAATATAGATAGGTTC	1078
Qy	904	AGATGTGACGCAATTTTACATTTCTACAAAGTTTCCATGCATGTATGTATGCTGAGGAC	963
Db	1079	GTAATGTGTAGGATTTTGTGATTCATTAATTCAGATCATCGATGATACAGAACAA	1138
Qy	964	GCAAAAAAGACTGTGTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTAAATGGGAA	1023
Db	1139	TTTAAGAAAGATTTTATTAAGTTCACATCAAAATGACAGG---CTAATGTTTAAAGTTAGA	1193
Qy	1024	AAACGATCCTAGAAATTTACTCCAAATTTGACATTTCTATACGCAAAAACAAGTTTAAACGAGT	1083
Db	1194	CGATGATCCTTGAATTACTAAATAATGAAAAATTTATTCG-AAAAACAAGCATAGATGAGT	1252
Qy	1084	TACCAAGATTATATATGTTTAAATTTGCGCATATGATGCTATGTTGTGTACACGTCCACTTA	1143
Db	1253	TGCCCAATTCATATATGTTTAAAGGTGATATAGTTAGTTAGTGAACAACGCTCCCA	1312
Qy	1144	CAGTTGATGAATTTGAAAAATTTTCTCGTGTCAAAAAGACAGATGAGTTTAAACGAG	1203
Db	1313	CAGTTGATGAATTTGAAAAAGTAAATTTTCAACGACAAACGACGCTTATGTTTAAACGAG	1372
Qy	1204	GGATTACAGCTCTCTGCGCAGGTTAGTGTGCTGTAATATACAGACTTTCAGCAGCGTAG	1263
Db	1373	GATACACTGCTTTGTGGCAGAAATTCCTGTTAGAAAATTAATTAATCTGATTTGATGAAATCG	1432
Qy	1264	TTCCGTTGCACTTACAGTACATTTGAAATTTGACATCTGTCAGATATTTAAATTTTAT	1323
Db	1433	TAAAGTTAGATGTTCAATATATCAATGATGATGCTATTTGGCTAGATATTAAGATTATTC	1492
Qy	1324	TAAAGACAGTGAAGTTGTTATTTGTTGAAGAGGGAAGTAATTAAGAT-----	1373
Db	1493	TCCTTAACATTAAAGGTAGCTTACTTGGGACAGAGACTTAAGTTAAAGGTAGAGTTGAAG	1552
Qy	1374	-----ATGAAAGTTGTGTTGTCGCTTCTTCAGGAGGACATTTGACTCTGATTT	1426
Db	1553	GAAATTAATGAATAATTTGTCTGTTGTTGTTCAAGGTGTGTATCTAGACACTTGAACCTT	1612
Qy	1427	GTTAAACCGTTTGAAGGAAGAACGTTTGGGTAACTTTGATTAAGAGATGC	1486

Db	1613	TTTGAAATCCATTGGGAAAAAGAAAGATGGTTTGGGTAACTCTTGATTAAGAAGATGC	1617
QY	1487	AAGAAGCTCTTTGGAAGATGAAAAATGTATCCATGTTACTCTTCCAAACAAATCGAATCT	1546
Db	1673	TAGAGGTAATTTGAAGAGAAAGATTTGTATATCATGTGCTCTTCCAAACCGTAAAGT	1733
QY	1547	CATTAAATTTAGTGAATACTTTCTTAGCTTCCAAATTTTACGTATGGAAGAACAGA	1606
Db	1733	CAAAACTGGGTAAAAATACATATTCTAGCTTTTBAAGTCCCTAGAAAAGAACACGA	1792
QY	1607	TGTTATTTATTCATCTGGTGGCGCGTGTCTGCCCTCTCTTTTACATCGAAAACTATT	1666
Db	1793	TGTTATATCATCATCTGGTGGCGGTGATGAGTACCATCTTTTATATTTGTAAGTTATT	1855
QY	1667	TGAGCAAGAAGACGATTATATTGAAGTATTGATCGAGTATATATCTTACATTACTGG	1722
Db	1853	TGGTTGTGAACCAATTTATATAGAGGTTTTCGACAGAGATATATAAACCAACTTTGACAGG	1912
QY	1727	AAACTAGTTTATCCCGTAACAGATTTTATATGTTACAGTGGGAAGAAATGAAGAAGT	1786
Db	1913	AAATTTAGTATCTGTACAGATTAATTTATTTATTTGTTACAGTGGGAAGAAATGAAAAAAGT	1972
QY	1787	ATATCCCTAAATCTATTAACTTGGGAGATTTTAAATGATTTTGTATACAGTAGAAGT	1844
Db	1973	TTATCTTAGGCAATTAAATTAGAGAGAAATTTTAAAGATTTTGTGCACGTGGGACA	2033
QY	1847	CATGAACAACAGTTTAATCGATTGATTAATAAGAGATGATTTATTGAAAAAAAATGGAAGT	1906
Db	2033	CATGAACAGCAGTTTCAACCGTCTTATTAAAGAGTGTATAGATTAAAAAGGACAGGTCT	2092
QY	1907	ATTAACCAAGAAATTTATTCAAAACAGATATTTCTGACTATATTCCGAATATTGCAAG	1966
Db	2093	ATTGATCAAGAGAGTTTCATTCAACCGGTTTACTCAGCTTTGAACCTCAGAAATTTCTAG	2152
QY	1967	TATTAATAATTTCTCAGTTTACAAAGAAATGGACAAATATATTAACAATCAGAAGTAGTT	2026
Db	2153	TGTCATAAATTTCTCTCATATGATGATATGAACTGTTCATGAGAGAAAGCTGAGATTGTT	2212
QY	2027	ATTGGCCACGAGAG 2040	
Db	2213	ATCACACATGCGCG 2226	
RESULT 12			
US-10-804-408-167			
; Sequence 167, Application US/10804408			
; Publication No. US20040253617A1			
; GENERAL INFORMATION:			
; APPLICANT: FANRONG, KONG			
; APPLICANT: GILBERT, GWENDOLYN			
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI			
; FILE REFERENCE: 675002-2001			
; CURRENT APPLICATION NUMBER: US/10/804,408			
; CURRENT FILING DATE: 2004-03-19			
; PRIOR APPLICATION NUMBER: PCT/AU02/01281			
; PRIOR FILING DATE: 2002-09-18			
; PRIOR APPLICATION NUMBER: AU PR 7749			
; PRIOR FILING DATE: 2001-09-19			
; NUMBER OF SEQ ID NOS: 182			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 167			
; LENGTH: 2226			
; TYPE: DNA			
; ORGANISM: Streptococcus agalactiae			
US-10-804-408-167			
Query Match 9.3%; Score 652.4; DB 8; Length 2226;			
Best Local Similarity 59.6%; Pred. No. 3.5e-82;			
Matches 1225; Conservative 0; Mismatches 791; Indels 38; Gaps 6;			
QY	16	TGGGATTAATTGATATGATAGCAAGTCAATTCCTGCAATCTTAAACAGTCAATPACCA	75
Db	182	TAGGAGTATTCACACAGTTGTATGTTATTTTCTGCAAGTTTACATTAACATTAATTA	241

QY 76 ATGCTGATTTAATG---TTCGGAATTTTATCATATGATGATGATTTTGCAT 132
Db 242 CTCCTCAATTTTAAAGCAATAAAGATTTATGTTGTTCTATGATACATATATGTC 301
QY 133 TTTTATATCTCGTATGCGAGTTGAATTTAGATATAGAGTATCTGATGAGTTGAAA 192
Db 302 TTTATCTTCTGATTTTACAGAGCTTTGGAGTGTGCTATCTGAAAGTTTAAAA 361
QY 193 AAACATTTACTATGATATATATTTTGCATTTTCTTACGGCAGATCATTTTGTGG 252
Db 362 TGGTATGAAATACAGCTTTTACTATATTTTCATATCAAGTTCATTTATTTTATTTA 421
QY 253 AGAATATTTCCGACTTTCAAGCGNGTGCCTGATTTTTCACATTAATAAATCTGTT 312
Db 422 AAAACCTTTTACACAGACACGCTTCTTTTATCTTTATGCTATGAAATTCGATTT 481
QY 313 TGGTATCTATTTAAGTAAATATTTAGCAGTTTAAGATAGCTTTCTATTTTCGACA 372
Db 482 TATGATCTATTTGAATTCATTTTAAATATATGAAATATCTTACGCTAAGTTT 541
QY 373 TCTATCAAAAAAGAGATTTTAAATTAACAACGCTGAACGATGGAAAAATATGCAAGTT 432
Db 542 CACGAGATACAAAGTTGTTTGTATACGATTAAGAGATCTTTATCAAAATGACCTTTA 601
QY 433 TATTTGAATCACTAAACAAATTCAAAAAATCTGTGCACTGTAGTTTAGTACAG 492
Db 602 GGAATTAAT--ACGACCATTAATTAATGCGCTGCTGATCTTGAGCTCCTCGAAAAAG 658
QY 493 AAATAGATAAATTAATTTATTCATTAACGCTCTATTAATCTGGAAGAGACTATAGAGT 552
Db 659 ATGTTATGATTTGAAACATACTCGTTAGAGATTAATAACAAAGATGCTTACTTACAG 718
QY 553 TTTCAACAAGGAGATGCTGACACAGCTCTTTTAAATCTACCAAGTGAATTTAG--- 609
Db 719 AGTTAACGTGCTTAACGTGTGATCAAGCTTTTATTAACATACCATTGAATTTTGTGA 778
QY 610 -----ACGTAAGCAATTCGTTTCAATTTTGAAGTTTGAAGTATGATGAAGCGTT 663
Db 779 AATACCAATTAACAAGATTAATTAATGACATTAAGCAATGGAGTGAATGTCATGTTA 838
QY 664 ATATTAATTCATTCGCTTTTACTGCTTGAACCAAAAAATCAACCTGAGTGAC 723
Db 839 ATGTAGAGCACTTATGCTTATGATTAATAGAGAAAAAGCAATTCACAACTTTTGAAGAT 898
QY 724 ATAGCATTTGATCTTTTTCACAAATTTTAAAGCTAGTCATATCATATGAAGAC 783
Db 899 ATAGTGTATTAATATTTATGAAATTTATATAATATAGTCACTTATAGCAAAACGAT 958
QY 784 TTTTGAATATACGCGAGCGGTAGTCCGGTTAATTTTGTGATATGTTTCTATTTTGT 843
Db 959 TTTTGAATATACGCGAGCGGTATTAATTAAGTTTGTCAATGTGGCAATTTTTC 1018
QY 844 TAGTTCAATTTATCTAGAGATGATGACCGGCTATTTTGTCAAGAAACGATTTGAC 903
Db 1019 TAGTTTCACAATACAAAAAGATGTGACCGGCTATTTTCTCAAAATAGATGATGTC 1078
QY 904 AGATGACGCGATTTTATACATTTTCAAGTTTCAATGATGATGATGATGATGATGATG 963
Db 1079 GTATGTAGAGATTTTATGATTTTAAATTCAGATCAATGAGATGATGATGATGATG 1138
QY 964 GCAAAAAAGACTTCTCAGCCCAAAACAGATGCAAGGTGGGTATGTTTAAATGGGAA 1023
Db 1139 TTTAAGAAAGATTTTATGTTTACATCAATATGACAGGG-----CTAATGTTAAGTTAGA 1193
QY 1024 AAAAGATCTTAGAATTTCTCAATTTGACATTTTCAAGCAAAAAACGATTTAGACGAGT 1083
Db 1194 AGATGATCTAGAAATTTACTTAATAATGAAAAATTTATTTG-AAAAACAAGCATATGATGAGT 1252
QY 1084 TACCAAGTTTATATGTTTAAATTTGCGATATGAGTCTAGTTGATACAGCTCCACCTA 1143
Db 1253 TGCTCATTTCTATATATGTTTAAAGGTGATATGAGTTTGTAGGAACAGCCCTCCCA 1312

QY 1144 CAGTGTAGAAATTTGAAAAATTAATCTCTGCTCAAAAAGAGCGATTTGATTTAAACGAG 1203
Db 1313 CAGTTAGTAAATATGAAAAAGATATTAATCAACGCAACAGCGCTTATGTTTAAAGCCAG 1372
QY 1204 GGAATTAACAGGCTCTGCGAGGTTAGTGTGCTAGTAAATATCAGACTTTCAGACGATG 1263
Db 1373 GAATCACTGGATTTTGTGGCAAAATATCTGTAGAAAAATATCATCTGATTTTGAATGAAATCG 1432
QY 1264 TTGCTGGAATTTAGCATATCATTTGATTAATTTGACTATCTGTCAGATATTAATTTTAT 1323
Db 1433 TAAAGTTAGATGTCATATATATCATGATGATGATGATGATGATGATGATGATGATGATG 1492
QY 1324 TAAAGACGTGAAGTGTATGTTTGAAGAGGGAAGTAAAGTAAAGT----- 1373
Db 1493 TCTTAACCTAAAGGATGCTTATCTTGGACAGGCTGCTAAAGTAAAGTAAAGTAAAG 1552
QY 1374 -----ATGAAGTTGTTGCTGCTTCTTCAAGGGGACATTTGACTCACTGATTT 1426
Db 1553 GAATATATGAAATTTGCTGCTTGTTCAGAGTGTGTCATCTAGCACACTTGAACCT 1612
QY 1427 GTTAAACCGTTTGGAGAGAGAAACGTTTGGGTAACATTTGATTAAGAGATGC 1486
Db 1613 TTTGAACCCATTTGGAGAAAAAGATAGGTTTGGGTAACCTTGAATAAGAGATGC 1672
QY 1487 AAGAGCTCTTTGAAGATGAAAAATGATTCATGTTTCTTCCAAAGAAATCGCAATCT 1546
Db 1673 TAGGATATTTCAAGAGAGATGATGATATGCTTCTTCCAAAGAAATCGTATAGT 1732
QY 1547 CATTAATTTAGGAAATTAATCTTCTTACCTTCAAAATTTTACGATGAGAGAAACGAGA 1606
Db 1733 CAAAACTTGTGTAATAATTAATTAATTTTACCTTTTAAGTCTTGAAGAAAGAAAGACAGA 1792
QY 1607 TGTATATTTGATCTGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1666
Db 1793 TGTATCATATCATCTGATGCGCTGATGACAGTACCATCTTTATATGTTAGTATAT 1852
QY 1667 TGAAGCAAGAGCATTTATTTAGATTTTGAATTTGATGAGTAAATTAATCTTACATTACTG 1726
Db 1853 TGGTGTAAAGCCGTTATATAGAGGTTTTCGACAGATTAAGTAAACAACTTTGACAGG 1912
QY 1727 AAAACATGTTATCCGTAACAGATTTTATTTGTCAGTGGAGAAATGAAGAGT 1786
Db 1913 AAAATTAAGTATCCGTAACAGATTAATTTATTTGTCAGTGGAGAAATGAAGAGT 1972
QY 1787 ATATCTTAATCTATTAATCTTGGAGATTTTATTTATGATTTTGTAAACATGAGAACT 1846
Db 1973 TTAATCTAAGCAATTAATTTAGAGAGATTTTATATATTTTGTACAGAGTGGGACA 2032
QY 1847 CATGAACAAGTTTAATGATGATTAAGATTAAGATTTTATTTGAAAAAAAATGCAAGT 1906
Db 2033 CATGAACAGAGTTCAACCGCTTATTAAGAGATGATTAAGTAAAGGAGACAGGTCT 2092
QY 1907 ATTAACGACGAATTTTATTAACAGGATATTTGACATATTTCCAGATTTTTCAGAG 1966
Db 2093 ATGATCAAAAGGTTTCAATCAAGGTTTACCTGACATTTGAACCTCGAATTTGACAG 2152
QY 1967 TATTAATAATTTCTCAGTTTCAAAAGAAATGGAACAATATATTAACAATCAGAGTAT 2026
Db 2153 TGGTCAAAATTTCTCTATATGATGATGATGATGATGATGATGATGATGATGATGATG 2212
QY 2027 ATTTGCAAGGAGG 2040
Db 2213 ATCACAATAGGCGG 2226

RESULT 13
US-10-804-408-162
; Sequence 162, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI

```

; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804.408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 162
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-162

```

Query Match	9.3%	Score 650.8	DB 8	Length 2217
Best Local Similarity	59.6%	Pred. No. 6e-82		
Matches 1224	Conservative 0	Mismatches 792	Indels 38	Gaps 6

QY	16	TGGCATTAATTGATATGATAGACAGTGGCAATTTCTGCAATCTTAA	CAAGTCATATACCAA	75
Db	173	TAGCGAGTATTCAAACMGTTGATGTTATTTTCTGCAAGTTTGACATTTAA	CATTATTA	232
QY	76	ATGCTGATTTTAATCG--TTCTGGAAATTTTATCATATGATGGTTCATTATTTGCAT	132	
Db	233	CTCCCAATTTTAAAGCAATAAAGATTATTTGTTGTTCTATGTATACATTATATGTTT	292	
QY	133	TTTTTATATCTCGTATGCCAGTCTGATTTGAGTATAGAGTAAATCTGATAGATTGAA	192	
Db	293	TTTATCTTTTCGATTTTATACAGACCTTTGGAGTCGGCTATCTTGAAAGTTTAAAA	352	
QY	193	AAACATTTAATAGTATAGTATTAATTTTGGAAATTTTCTTACGGGATACATTTTGTGG	252	
Db	353	TGCTATTTGAAATVACAGCTTTACTATATTTTCAATATACAGTTTCATTATTTTATTTTAA	412	
QY	253	AGATTAATTTTGCACCTTCAAGACGCTGGTGCCTGATTTTGCATTAATTAACCTGTT	312	
Db	413	AAAACCTTTTACACAGCACGACCTTCTCTTTTACTTTATTTGCTATGTAATTCGATTT	472	
QY	313	TGCTATACCTATTTAAGCTAATTTAAGACGTTTAAAGATAGCTTCTATTTTTCGACAA	372	
Db	473	TATATATCTATGTAATTCATTTTAAATATTTATCGAAATATTTCTTAACGTAAGTTT	532	
QY	373	TCTATCAAAAAAGACGATCTTAATTACACGGCTGAAACGATGGAAAAATATGCAAGTT	432	
Db	533	CACGAGATACCAAGTTGTTTGGATTAACGAATTAAGGATCTTTATCAAAAAATGACCTTAA	592	
QY	433	TATTTGATATCAATTAACAAATTCAAAAAATCTGTTGCATGTGTAATTTTAGSTACAG	492	
Db	593	GGAAATAAAT--ACGACCATAATTATATATGCTGCTGTATCTTGACCTCTGAAAAAG	649	
QY	493	AAATAGATTAATAATTTATCATTTACCTACCGCTCTATTAATTTCTGGAAAGACATATAGAGT	552	
Db	650	ATTGTTATGATTTGAACAATACTACCTGTTAAAGATATATAAACAAGATGCTCTTACTTAG	709	
QY	553	TTTCAACAAGGGAAGTGTGACCAACGCTTTATTAATCTTACCAAGTAGTTTTAG--	609	
Db	710	AGTTAACCTGCTTAAGTGTGATCAAGCTTTTATTAATCATACCACTGAATTAATTTGGTA	769	
QY	610	-----ACGTTAAGCAATTCGTTTCAGATTTTGAGTTGTAGTATTTGATGTAAAGCTTG	663	
Db	770	AATACCAATTCAGATATTAATTTATATGACATTTGAAGCAATGGAGTGAATGTCAATGTTA	829	
QY	664	ATATTAAATTCATGCGGTTTTTACCTGCGTGGAAAAACAANAATATCCATCTGAGGTGACC	723	
Db	830	ATGTAGAGGACCTTAGCTTTGATATATATAGAGAAAGCAATCCAACTTTGAAAGAT	889	
QY	724	ATAGCATTTGTAATCTTTTCCAAATTTTATTAAGCTTAGTCAATATCATATGATGAACGAC	783	
Db	890	ATTAGGTATTAACATATTTCTATGAATAATCTTAATAATATAGTCACTTAATAGCAAAAACGAT	949	
QY	784	TTTTTGATATATCTGGAGCGGTAGTCGGGTTAATTAATTTGTGTATAGTTTCTATTTTGT	843	

Db	950	TTTTGATATACAGCGGTCTATTATAGTGTGCTCATATGCGGATGTGGCAATTTTC	1009
Qy	844	TAGTTCCAATTTATCTGATGAGATGCTGACCGGCTATTTTGGCTCAGAAAACGATTTGAC	903
Db	1010	TAGTTTCGCAAAATCAGAAAAAGATGCTGGACCGGCTATCTTTTCTCAAAAATAGATGAGTTC	1069
Qy	904	AGATGCAACCATTTATACATTTCAACAAGTTTGCATGATGTATGTGATGCTGAGGAGC	963
Db	1070	GTATGCTTAGATTTTATGATTTCTATTAATTCAGATTAATGCGAGTATGATGCGAGACAAA	1129
Qy	964	GCAAAAAAGACTTGGCTCAGCCAAAACAGATGCAAGGGTGGTATGTTTTAAATGGGAA	1023
Db	1130	TTTAGAAAGATTTATTTAGTTTCAACATTAATGACAGG-----CTAATGTTTAAGTTAGA	1184
Qy	1024	AAACGATCTTAGAATTACTCCAAATTTGACATTTCAATCGCAAAAACAAGTTTGAACGAT	1083
Db	1185	AGATGATCTTAGAATTTCTAAAAATAGGAAAATTTATTCG-AAAAACAGATGATGATGAGT	1243
Qy	1084	TACCAAGTTTTATTAAGTTTAAATTTGGCGATATAGTCTAGTTGGTACACGTCCACTA	1143
Db	1244	TGCTCAATTCATAAAGTTTAAAGGTATATGATTTAGTAGGAACACGGCTTCCCA	1303
Qy	1144	CAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAAGACAGATTTGATTTAAACAG	1203
Db	1304	CAGTTGATGAATATGAAAAATATATTTCAACGCAAGACGACCGCTTAGTTTAAAGCAG	1363
Qy	1204	GGATTTACAGGTTCTCTGGCAGGTTAGTGTGCTGATTAATACAGACTTGCACGACGTAG	1263
Db	1364	GAATCACCTGGTTGTGGCAAAATATCTGTAGAAATATATTTACTGATTTTGAATGAAATCG	1423
Qy	1264	TTGGTTGGACTTAGCATACATTGATTAATTGGACTATCTGGTCAGATATTTAAATTTTAT	1323
Db	1424	TAAAGTTAGATGTTCAATATATCAATTAAGGTCATATTTGGTCAGATTAATTAAGATTATTC	1483
Qy	1324	TAAAGCAGGAAAGTTGTATTGTTGAGACAGGGAAGTAGTAAGTAAAGTAT-----	1373
Db	1484	TCTTACACTTAAGAGTATGTTTACTCGGGACAGAGCTAAGTAAAGTTAAGTTTGAAG	1543
Qy	1374	-----ATGAAAGTTTGTGTGTCGGTCTTCAAGGGGACATTTGACTCATTTGATATT	1426
Db	1544	GAATATATAGAAAAATTTGTCTGGTGGTTCAAGTGGTGCATCTAGCACACTTGAACCT	1603
Qy	1427	GTTTAAACCGTTTGGAGAAGAAAGAAAGCTTTTGGGTAACTTTGATTAAGAAGATGC	1486
Db	1604	TTTGAAAACCAATTTGGGAAAAAGAAAGATAGTTTTGGGTAACTTTGATTAAGAAGATGC	1663
Qy	1487	AAGAAGCTTTTGAAGAAATGAAAAAATGATTCATGATTACTTTCCAACAATGCAATCT	1546
Db	1664	TAGAGATATCTTAGAGAGAGAGATGTATATCACTGTCTTTTCCAAACAACCGTATATGT	1723
Qy	1547	CATTAAATTTAGTGAATAATATCTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACGAGA	1606
Db	1724	CAAAAACTTGGTAAAAAATATCTATTTCTAGCTTTTAAAGTCTTAGAAAAAAGAACGAGA	1783
Qy	1607	TGTTATTTATTTCACTCTGGTCGGCCGCTGTCTGCCCTTTTCAATCGAAAACTATT	1666
Db	1784	TGTTATCATATCATCTGGTCCGCCCTGACAGATCACAATCTTTTAAATTTGGTAAAGTTATT	1843
Qy	1667	TGAGAGCAAAACGATTTATATTAGAGATTTGATTCAGAGTAAATAATCTATTAATCTGG	1726
Db	1844	TGGTTGTAAACCGTTTATATTAATTAAGGTTTTCAGACAGATATGATTAACCAACTTTGACAGG	1903
Qy	1727	AAAACATGTTTATCCGCTAACAGATATTTTATTTGTTCACTGCGGAAGAAATGAGAAGGT	1786
Db	1904	AAAATTTAGTATATCTGTACACATTAATTTATTTGTTCACTGCGGAAGAAATGAAAAAGT	1963
Qy	1787	ATATCTTAATCTATTAACCTTGGGAGTATTTTATATGATTTTGTAGACATGACACT	1846
Db	1964	TTATCTTAAGCAATTAATTTAGAGAGAAATTTTATATGATTTTGTGCACAGTGGGAGCA	2023
Qy	1847	CATGAACAAACAGTTTATGCAATTTGATTAAGATGATTTATTTGAAAAAATATGAGAT	1906
Db	2024	CATGAACAGCAATTTCAACCGCTTTATTAAGAAAGTTGATGATTTAAAGGACAGCGTCT	2083

QY	1907	ATAA	CCGACGAATATTATTATTC	CAACAGATATTCGACTATATTC	CAAGATATTCGAG	1966
Db	2084	ATTGAT	CAGAAGTGTTCA	TCCAAACGGGTACTCAGACTTCGAACCTC	CAAAATTTGTCAG	2143
QY	1967	TATATAAAATTTTCAG	TTCACAAAGAAATGGA	CAATATTATTA	CAAAATCAGAGTAGTT	2026
Db	2144	TGTC	CAAAATTTCTCTCATATGATGATATGAACTT	TACAGAAAGAAAGCTG	AGATTGTT	2203
QY	2027	ATTG	CCACGAGAG	2040		
Db	2204	ATCA	CAATGAGCGG	2217		

```

RESULT 14
US-10-804-408-173
/ Sequence 173, Application US/10804408
/ Publication No. US20040233617A1
/ GENERAL INFORMATION:
/ APPLICANT: FANRONG, KONG
/ APPLICANT: GILBERT, FANRONG, KONG
/ TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
/ FILE REFERENCE: 675002-2001
/ CURRENT APPLICATION NUMBER: US/10/804,408
/ CURRENT FILING DATE: 2004-03-19
/ PRIOR APPLICATION NUMBER: PCT/JP02/01281
/ PRIOR FILING DATE: 2002-09-18
/ PRIOR APPLICATION NUMBER: AU PR 774,9
/ PRIOR FILING DATE: 2001-09-19
/ NUMBER OF SEQ ID NOS: 182
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 173
/ LENGTH: 2226
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Consensus sequence
US-10-804-408-173

```

Query Match	9.3%;	Score 649.2;	DB 8;	length 2226;
Best Local Similarity	59.5%;	Pred. No. 1e-81;		

Oy	16	TGGCATTATTTGATATGATAGACAGTTGCAATTTCTGCAATCTTAAACAGCTCATATACCA	75
Db	182	TAGCGATGATCAAAACAGTTGGTATTATTTTTCGCAAGTTGACATTTACATTAATTA	241
Oy	76	ATGCTGATTTAAATCG--TTCTGGAATTTTATCATTAATGATNGTTCATTTATTTGCAT	132
Db	242	CTCCCAATTTTAAAGCATAAAGATTTATGTGTTGTTCTATTTGATACATTTATATGTCT	301
Oy	133	TTTTTATATCTGGATGACAGTTGAATTTGAGATAGAGTATCTGATAGATTGAAA	192
Db	302	TTTATCTTTCTGATTTTACAGAGACTTTTGAGTGTGGCTATCTTGAAGATTAA	361
Oy	193	AAACATTACTATAGTATATATATATTTTGCATTTTCTTACGGCAGTATCATTTTGGTTGG	252
Db	362	TGGTATTTGAATTAACGCTTTTACTATATTTTCAATTCATCAAGTTCATTAATTTTATTTTAA	421
Oy	253	AGATTAATTTGCACTTTCAAGCGTGGTGCCTGATTTTCACATTAAATPAACTTGCTTT	312
Db	422	AAACCTTTTACAAAGCAGCAGCTTCCCTTTTATCTTTTATTTGCTATGAATTCGATTT	481
Oy	313	TGGTATTACTATTAAACGTAAATTTATPAGAGTTTAAAGATAGCTTTCTATTTTGGACA	372
Db	482	TATTTATATCTATTTGAATTCATTTTAAATATTATGAAAATATTTCTTACGCTAACTTTT	541
Oy	373	TCTATCAAAAAAAGACGATCTTAATTACAACGGCTGAACGATGGAAAAATATGCAAGTTT	432
Db	542	CACGAGATACCAAGTGTGTTTGTATTAACGAATAGGATTTCTTTATCAAAAATGACCTTTA	601
Oy	433	TATTTGATTCATTAACCAATTCAAAAAATCTTGTGCATGTGATTTTNGTACAG	492

Db	602	GGAAATAAAT---	ACGACCAATAATTATATACGCTGCTGTATCTTGGACATCTCTGAAAAG	655
Oy	493	AAATGATTAATAATTATTAATCATTAACCGCTCTATTAATTCGTGGAAGAGCTATAGACT	552	
Db	659	ATTGTTATGATTTTGAAACATPAACCTGTTAAGGATATATAAACAAATGCTCTTACTTCAG	718	
Oy	553	TTTCAACGAAGAGAGTGGTCCAGCACGCTCTTATATAATCTACCAAGAGTTTTAG---	609	
Db	719	AGTTAAACCTGCTTAACTGTGTATCAAGCTTTTATTAACATACCCTATGATATATTTGGTA	778	
Oy	610	-----ACGTAAGCAATTCGTTTCAGATTTTGAGTTGTGAGTATTCGATGTAGACGTTG	663	
Db	779	AATACCAATATACAGATATATTAATATGACATTGGAAGCAATGGAGATGTGTCAATGTTA	838	
Oy	664	ATATTAAATTCATTCGGTTTTCACGCGTTGAAAAACAAAAAAATCCAATGCTAGGTGACC	723	
Db	839	ATGTAGAGGCACTTAGCTTTGATTAATATATGAGAAAAAGCAATCCAAATTTTGAAGAT	898	
Oy	724	ATAGCATGTAACTTTTCCACAAATTTTATTAAGCTTAGTCATATCATGATGAAGAACG	783	
Db	899	ATAGGTTATTAACAATATCTATGAATAATCTATTAATAATATAGTACCTTATAGCAAAACGAT	958	
Oy	784	TTTTGGATTACTCCGAGCGGATGCGGGTTAATATTGTGGTATGTTCTAATTTGT	843	
Db	959	TTTTGGATATCAACGGGTGCTATTAATAGTTTGCTCATATGTGGCANTGTGGCAATTTTTC	1011	
Oy	844	TAGTTCCAAATTAATCGTAGAGATGTGACCGGCTAATTTTGTCTCAGAAACGATTTGAC	903	
Db	1019	TAGTTCCGCAATTCAGAAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGCTC	1077	
Oy	904	AGATATGACGCATTTTACATTCATTCACAGTTTCGATTCGATGTATGTATGTATGCTGAGAGAC	963	
Db	1079	GTAATGCTGAGATTTTATGATTCATTAATTTAGATCAATGCGAGTAGATGCAACAA	1133	
Oy	964	GCAAAAAACACTGCTCAGCCAAAACCAAGATCAAGGGTGGGATGTTTAAATGGGAA	1022	
Db	1139	TTAAGAAAGATTTATTAATGTTCAACATCAATATGACAGG-----CTAATGTTTAAGTTAGA	1199	
Oy	1024	AAACGATCCTAGAAATTAATCAATTTGGACATTTCAATAGCAAAAAACAAGTTTAGACGAGT	1088	
Db	1194	CGATGATCCTAGAAATTAATTAATAATGAGAAATTTATTCG--AAATAACAGATATGATGAGT	1255	
Oy	1084	TACACAGTTTATTAATGTTTAAATTTGGGATATGAGTCTAGTTGTACACGCTCACTTA	1144	
Db	1253	TGCTCTCAATTCATATATGTTTAAAAAGGATATGAGTTTATGATAGAGAACCGCTTCCCA	1311	
Oy	1144	CAGTTGATGAATTTGAAAAATATATCTCGTGTCAAAGAAGACGATGAGTTTAAACAG	1201	
Db	1313	CAGTTGATGATATGAAAAAGTATATTAATTCACACGCAAGACGACCCCTTAGTTTAAAGCAG	1377	
Oy	1204	GGAATACAGGTCTCTGCGACGTTAGTGTGTGATATATCACAGACTTGACGACGTAG	1266	
Db	1373	GATTCACGTGTTTGTGGCAAAATATCTGTATGAATATATTAATCTGATTTTGAAGAATCG	1433	
Oy	1264	TTGCGTTGCACTTAGCATATCATTAATTTGGACTATCTGATCGATATTAATAATTTAT	1322	
Db	1433	TAAAGTTAGATGTTCAATATATCAATGAATGCTATTTGGTCAAGATATTAAGATTAATTC	1499	
Oy	1324	TAAAGCAGGAAGATGTTATTTGAGAGAGGGAAGTAAAGTAT-----	1377	
Db	1493	TCTTAACACTAAAGGTATGTTTACTTGAGCAGAGACTAAGTAAAGTTAGGTTGAAAG	1558	
Oy	1374	-----ATGAAAGTTTGTGTGTCGGTTCTTCAGGGGACATTTGACTCATTTGATTT	1427	
Db	1553	GAAATATATTAATAATTTGTCTGTGGTTTCAAGTGGTGCTCATCTAGCAACATTTGAACCT	1611	
Oy	1427	GTTTAAACCGTTTGGAGAAGAGACGTTTTTGGGTAAACATTTGATTAAGAGATGC	1486	
Db	1613	TTTGAACCCATTTTGGGAAAAAGAGATAGGTTTGGGTAACTTTGATTAAGAGATGC	1679	
Oy	1487	AAGAAGCTTTGAAGAAATGAATAAAATGATATTCATCTTTTCAACAATGCAATCT	1545	
Db	1673	TAGAGATATCTATAGAGAGAGATGTATATTCATTCCTTTTCCAAACAAACGTAATGT	1737	

QY 1547 CATTAATTTAGTAAAAATCTTCTTACCTTCAAAATTTTACGTGATGAAACAGA 1606
 |||||
 Db 1733 CAAAACTGGTAAAAAATCTATCTAGCTTTTAAAGTCTCTTAAAGAAAGAACAGA 1792
 |||||
 QY 1607 TGTATATTTTCACTGTCGGCGCGTGTCTGCCCTTCTTTTACATCGAAAACTATT 1666
 |||||
 Db 1793 TGTATATCATCATCTGTCGGCGCGTGTGACAGTACATCTCTTTTATATGTGTAAGTATT 1852
 |||||
 QY 1667 TGGAGCAAAACATTTATATTTGAAGTATTTGATTCAGTATTAATTAATCTAATTAAGT 1726
 |||||
 Db 1853 TGGTGTAGAACCGTTTATATAGAGTTCGACAGAGTATTAACCACTTTGACAGG 1912
 |||||
 QY 1727 AAAACTAGTTTATCCGCTAACAGATTTTATTTTGTTCAGTGGGAGAAATGAAGAGT 1786
 |||||
 Db 1913 AAAATTTAGTATCTGTTAAACAGATTAATTTATTTTGTTCAGTGGGAGAAATGAAGAGT 1972
 |||||
 QY 1787 ATATCTAAATCTATTAACCTTGGGAGTATTTTAAATTTTGTATACAGTAAAGT 1846
 |||||
 Db 1973 TTTATCTTAAGCAATTAATTTAGAGAAATTTTATTTTGTTCAGTGGGAGCA 2032
 |||||
 QY 1847 CAGAACACAGTTTATTAATGATTAAGAGATTTATTTGAAAAAATGAAGT 1906
 |||||
 Db 2033 CAGTAAACAGAGTCAACCGTCTTATTAAGAGTGTATTAAGAGGACAGGTGT 2092
 |||||
 QY 1907 ATACCGAAGAAATTTATTTCAACAGAGATTTCTGACTATATTCAGAAATTTGCAAG 1966
 |||||
 Db 2093 ATGATCAACAGAGTTCATTAACAGGAGTTCACAGCTTGAACCTCAAGATTTGAC 2152
 |||||
 QY 1967 TATATAAAATTTCTCAGTTTCAAAAGAAATGAACATATATTAACAATGAGAGT 2026
 |||||
 Db 2153 TGGTCAAAATTTCTCTCATATGATGATATGAACCTTACATGAAGAGAGTGTGT 2212
 |||||
 QY 2027 ATTTGCCAGGAGG 2040
 |||||
 Db 2213 ATCACACATGCGG 2226
 |||||

RESULT 15

US-10-804-408-171
 ; Sequence 171, Application US/10804408
 ; Publication No. US20040253617A1
 ; GENERAL INFORMATION:
 ; APPLICANT: FANRONG, KONG
 ; APPLICANT: GILBERT, GRENDOLYN
 ; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
 ; FILE REFERENCE: 675002-2001
 ; CURRENT APPLICATION NUMBER: US/10/804,408
 ; PRIOR FILING DATE: 2004-03-19
 ; PRIOR APPLICATION NUMBER: PCT/AU02/01281
 ; PRIOR FILING DATE: 2002-09-18
 ; PRIOR APPLICATION NUMBER: AU PR 7749
 ; NUMBER OF SEQ ID NOS: 182
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 171
 ; LENGTH: 2226
 ; TYPE: DNA
 ; ORGANISM: Streptococcus agalactiae
 US-10-804-408-171

Query Match 9.3%; Score 648.8; DB 8; Length 2226;
 Best Local Similarity 59.5%; Pred. No. 1.1e-81;
 Matches 1222; Conservative 1; Mismatches 793; Indels 38; Gaps 6;
 QY 16 TGGCATATTTGATGATAGCAGTTTCTGCAATCTTAACAAGTCATATACCA 75
 |||||
 Db 182 TAGCATGATTTCAACAGTGTGTGTTATTTTCTGCAATTTGACATTAACATTA 241
 |||||
 QY 76 ATGCTGATTTAAATCG--TTCTGGAATTTTATCATTAATGATGTTCAATTTTGCAT 132
 |||||
 Db 242 CTCCTCAATTTTAAAGCAATTAAGATTTATTTGTTGTTCTATGATCATTAATATGTCT 301
 |||||

QY 133 TTTTATATCTCGTATGCCAGTTGAATTTGATATAGAGTTAATCTGATAGAGTTGAAA 192
 |||||
 Db 302 TTTATCTTTTGTATTTTATACAGAGACTTTTGAGTGTGGCTATCTTGAAGAGTTTAAAA 361
 |||||
 QY 193 AAACATTTAATCTATGATATATTTTGAATTTTCTTACAGGAGATCATTTTGTGTG 252
 |||||
 Db 362 TGGTATTAAGAAATACCTTTTACTATATTTTCAATCAAGTTATATTTTATTTTAA 421
 |||||
 QY 253 AGAATTAATTTGCGACTTTTCAAGACGTGTGCGCTGTATTTTCAATTAATTAACCTGTT 312
 |||||
 Db 422 AAAACTTTTATACAGACACGACTTTCTTTTACTTTTATTTGCTATGAATTTGATTT 481
 |||||
 QY 313 TGGTATATCTATTTAAAGTATTTATTAAGAGCTTTAAGATAGCTTTCTATTTTGA 372
 |||||
 Db 482 TATTATATCTATTAATGATTTTAAATATTTAAGAAATATCTTACGTAAGTTT 541
 |||||
 QY 373 TCTATCAAAAAAGAGATTTCTAATTAACAAGCTGACAGATGGGAAATATGCAAGTT 432
 |||||
 Db 542 CACAGATACCAAGTGTGTGTATGAATGAATGAAGATTTCTTATCAAAAATGACCTTTA 601
 |||||
 QY 433 TATTGAATGACATTAACAAATTCAAAAAATCTGTGCAATTTGATTTAGTGTAG 492
 |||||
 Db 602 GGAATTAAT--ACGACCATTAATTAATGCGCTGTGTATCTGTGATTTCTGAAAAG 658
 |||||
 QY 493 AAATAGATTAATAATTTAATTTATCATTAACGCTTATTTATCTGGAAGAACTATAGT 552
 |||||
 Db 659 ATTTGATATGATTTGAAACATTAACCTGTTAAGATTAATTAACAAAGATCTTACTCAG 718
 |||||
 QY 553 TTTCAACAAAGAGAGTGTGACACAGCTTTTAAATCTACCAAGTGAATTTAG-- 609
 |||||
 Db 719 AGTTAACTGCTTAACGTTGATCAAGCTTTTATTAACATTAACCATTAAGTAAATTTG 778
 |||||
 QY 610 -----ACGTAAGCAATTCGTTCAATTTTGAATTTGATTTAGATTTAGTAAGCTGTG 663
 |||||
 Db 779 AATACCAAAATCAAGATTTATTAATGAATTAAGCAATGAAGAGTATGTAATTTTA 838
 |||||
 QY 664 ATATTAATTCATTCGCTTTTACTGCGTTGAAAAAACAATACTGCTAGTGAGC 723
 |||||
 Db 839 ATGTAGAGCACTTACTTTGATTAATTAAGAGAAAGCAATTCGAACTTTGAAGAT 898
 |||||
 QY 724 ATAGATTTGTAATTTTTCACAAATTTTAAAGCTTACTATATCATGATGAACGAC 783
 |||||
 Db 899 ATAGTATTTATCATATTTCTATGAATTTCTATTAATTAAGTCACTTATAGCAAAACGAT 958
 |||||
 QY 784 TTTTGAATATCTCGAGCGGTAGTGGGTTATTTTGTGTATAGTTCATTTTGT 843
 |||||
 Db 959 TTTTGAATATCAAGGAGTGTATTAATGAGTTTGTCTATGTCATTTGCGCAATTTTTC 1018
 |||||
 QY 844 TAGTCCAAATTTATTCGATGATGAGACCGGCTATTTTGTCTCAGAAACGAGTTGAC 903
 |||||
 Db 1019 TAGTTCACAAATCAAAAAAGATGTGACCGGCTATTTTCTCAGAAATGAGTAGTC 1078
 |||||
 QY 904 AGATGAGCGCATATTTTACATTTTCAAGATTTGATGATGATGATGATGATGATGATG 963
 |||||
 Db 1079 GTATGTGTAGATTTTATGATTTCTATTAATTTGATGATGATGATGATGATGATGATG 1138
 |||||
 QY 964 GCAAAAAAGACTTGTCTCAGCAAAACAGATGCAAGGTGGGTATGTTTAAATGGGAA 1023
 |||||
 Db 1139 TTAAAGAAAGATTTATAGTTTCAATCAATGACAGG-----CTATATGTTAAGTTAGA 1193
 |||||
 QY 1024 AAAGATCTAGAAATTTACTCCATTTGAGCACTTTCAATGCAAAACAAAGTTGAGCAGT 1083
 |||||
 Db 1194 CGATGATCTAGATTTACTTAATTAATGAGAAATTTATTTCC-AAAACAAAGATGATGATG 1252
 |||||
 QY 1084 TACCAAGATTTTAAATGTTTAAATTTGCGATATGAGTCTAGTTGATCAAGTCACTTA 1143
 |||||
 Db 1253 TGCCTCAATTTCTATATGTTTAAAGGATGATGATTTAGAGTAAACAGCGCCCTCCA 1312
 |||||
 QY 1144 CAGTATGATTTGAAAAATATATCTCTGTGTCAAAAAGAGCAATTTGATTTAAACAG 1203
 |||||
 Db 1313 CAGTTATGAAATATGAAAGATTAATTAACGAGAGGACGCTTATGTTTAAAGCAG 1372
 |||||
 QY 1204 GATTAACAGGTCTCTGCGAGGTTAGTGTGTGTATGATTAATCAAGACTTGAAGAGTAG 1263
 |||||

Db 1373 GAATCACTGGTTGGCGAAATATCGTAGAAATATATCTGATTTTGTATGAAATCG 1432
QY 1264 TTCCGTTGGACTTAGATACATTGATAATTGACTATCTGGTCAGATATTAATAATTAT 1323
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGGCTATTTGGTCAGATATTAAGATTATTC 1492
QY 1324 TAAAGCAGTGAAGTTGATTTGTTGAGAGAGGAGTAAGTAAAGTAT----- 1373
Db 1493 TCCTAAGACTTAAGTAGTTTACTCGGACAGAGACTAAGTAAAGTAAAGTTGAAG 1552
QY 1374 -----ATGAAAGTTGTTGGTCTGCTTCAGGGGACATTGACTCATCTGTATTT 1426
Db 1553 GAATATATGAATAATTGTCGTGGTTCAAGTGGTGGTCATCTAGCACACTTGAACT 1612
QY 1427 GTTAAACCGTTTGGAGGAAGAAAGCGTTTGGGTAACTTGTATGATTAAGAGATGC 1486
Db 1613 TTTGAAACCCATTGGGAAAGAAAGATAGTTTGGGTAACTTGTATTAAGAGATGC 1672
QY 1487 AAGAACTCTTTGAGAAATGAATAAATGATCCATGTTACTTTCACAAATCGCAATCT 1546
Db 1673 TAGAGTATTTCAAGAGAGAGATGATATCATTTGTTCTTCCAAACAAACGTATATG 1732
QY 1547 CATTAATTTAGTAAAAATACCTTCTAGCTTCAAAATTTTACGTGATGAGAAACACA 1606
Db 1733 CAAAAAATTGGTAAAAAATCTATCTAGCTTTAAGTCTTGAAGAAAGAAAGACACA 1792
QY 1607 TGTATATTTTCACTCGTGGGCGCGTGGTCCCTTCTTACATGGGAAACATTT 1666
Db 1793 TGTATATATATCATCTGGTCCGCTGAGACATCTTCTTATATTTGTTATTT 1852
QY 1667 TGGAGCAAGACGATTTATTTGAAGATTTGATCGAGTTAATTAATCTACATTAATCG 1726
Db 1853 TGGTTGTAGACCGTTTACATAGAGTTTTCGACAGATGATTAACCACTTTGACAG 1912
QY 1727 AAAACTAGTTATCCGTTACAGATTTTATTTTCACTGGAGAAATGAAGAGT 1786
Db 1913 AAAATTAGTATCTGTGTACAGTAAATTTATTTGTTCACTGGAGAAATGAAGAGT 1972
QY 1787 ATATCTTAATCTATTAATCTTGGGAGTATTTTATATGATTTTGTAACTAGAGAACT 1846
Db 1973 TTTATCTTAAGCAATTAATTTAGAGGAATTTTAAATTTTGTACAGTGGGACA 2032
QY 1847 CATGAACCAAGTTTATCGATTGATTAAGAGATGATTTATTTGAAAAAATGGAAGT 1906
Db 2033 CATGAACAGAGTTCAACCGTCTTATTAAGAGATGATTAAGAGAGAGAGTCT 2092
QY 1907 ATAAACGACGAAATATTTATCAACAGATATTTGACTATATTTCCAGAAATTTGCAAG 1966
Db 2093 ATTGATCAAGAGTTCATTTCAACAGGTTACTCAGACTTTGAACCTCAGAAATTTGCAAG 2152
QY 1967 TATTAATAAATTTCTCACTTCAAGAAATGGAACATATATTAACAATCAAGAGTAGTT 2026
Db 2153 TGGTCAAAATTTCTCTCATATGATATGAACCTTATCATGAAGAGAGCTGAGATTGTT 2212
QY 2027 ATTTGCCAGGAG 2040
Db 2213 ATCAGACATGCGG 2226

Search completed: December 25, 2005, 17:21:03
Job time : 3377 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 16:11:36 ; Search time 22524 Seconds
(without alignments)
17645.596 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgcaacgaattgca.....gcaaacgcaaaagaagctt 6992

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 5883141 seqs, 28421725653 residues

Word size: 0

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_fm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_scs:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hrg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6992	100.0	6992	1	AF155804 Streptococcus
2	611	8.7	15401	1	AF118389 Streptococcus
3	28	0.4	17468	1	AY057915 Streptococcus
4	28	0.4	17468	6	AX283715 Streptococcus
5	27	0.4	16032	6	CQ963297 Streptococcus
6	27	0.4	19798	1	CR931678 Streptococcus
7	27	0.4	21562	1	STR289861 Streptococcus
8	27	0.4	110000	1	CP000023_09 Streptococcus
9	27	0.4	165328	9	AC121782 Streptococcus
10	27	0.4	183094	9	AC087903 Streptococcus
11	27	0.4	224451	9	AC087062 Streptococcus
12	26	0.4	447	6	CQ645816 Streptococcus
13	26	0.4	450	6	AX607685 Streptococcus
14	26	0.4	688	1	AY257676 Streptococcus
15	26	0.4	688	1	AY257677 Streptococcus
16	26	0.4	688	1	AY257678 Streptococcus
17	26	0.4	688	1	AY257679 Streptococcus
18	26	0.4	688	1	AY257680 Streptococcus

19	26	0.4	688	1	AY257681 Streptococcus
20	26	0.4	688	1	AY257682 Streptococcus
21	26	0.4	688	1	AY257683 Streptococcus
22	26	0.4	688	1	AY257684 Streptococcus
23	26	0.4	688	1	AY257685 Streptococcus
24	26	0.4	2217	1	AF332903 Streptococcus
25	26	0.4	2217	1	AF332904 Streptococcus
26	26	0.4	2217	1	AF363038 Streptococcus
27	26	0.4	2217	1	AF363039 Streptococcus
28	26	0.4	2217	1	AF363040 Streptococcus
29	26	0.4	2217	1	AF363045 Streptococcus
30	26	0.4	2217	1	AF363050 Streptococcus
31	26	0.4	2217	1	AF363051 Streptococcus
32	26	0.4	2217	1	AF363052 Streptococcus
33	26	0.4	2217	1	AF367973 Streptococcus
34	26	0.4	2217	1	AF381031 Streptococcus
35	26	0.4	2220	1	AF332912 Streptococcus
36	26	0.4	2220	1	AF363035 Streptococcus
37	26	0.4	2226	1	AF332905 Streptococcus
38	26	0.4	2226	1	AF332906 Streptococcus
39	26	0.4	2226	1	AF332907 Streptococcus
40	26	0.4	2226	1	AF332908 Streptococcus
41	26	0.4	2226	1	AF332909 Streptococcus
42	26	0.4	2226	1	AF332910 Streptococcus
43	26	0.4	2226	1	AF332911 Streptococcus
44	26	0.4	2226	1	AF332913 Streptococcus
45	26	0.4	2226	1	AF363032 Streptococcus

ALIGNMENTS

RESULT 1
LOCUS AF155804
DEFINITION Streptococcus suis strain 655 CpsII (cpsII) gene, partial cds;
ACCSSION AF155804
VERSION AF155804.1 GI:6601338
KEYWORDS Streptococcus suis
SOURCE Streptococcus suis
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 6992)
AUTHORS Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.
TITLE Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor
JOURNAL Infect. Immun. 67 (4), 1750-1756 (1999)
PUBMED 10085014
REFERENCE 2 (bases 1 to 6992)
AUTHORS Smith,H.E., Veenbergen,V., van der Velde,J., Damman,M., Wisselink,H.J. and Smits,M.A.
TITLE The cps genes of Streptococcus suis serotypes 1, 2, and 9: development of rapid serotype-specific PCR assays
JOURNAL J. Clin. Microbiol. 37 (10), 3146-3152 (1999)
PUBMED 10488168
REFERENCE 3 (bases 1 to 6992)
AUTHORS Smith,H.E., Veenbergen,V., Van der Velde,J., Damman,M., Wisselink,H.J. and Smits,M.A.
TITLE Direct Subtyping
JOURNAL Submitted (02-JUN-1999) Bacteriology, Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The Netherlands
FEATURES
source 1..6992
/organism="Streptococcus suis"
/mol_type="genomic DNA"
/strain="655"
/serotype="1"

```
gene
CDS
    /db_xref="taxon:1307"
    <1..1367
    /gene="cpsA1"
    <1..1367
    /note="putative glycosyltransferase"
    /codon_start=3
    /transl_table=11
    /product="CpsA1"
    /protein_id="AAF18940.1"
    /db_xref="GI:6601339"
    /translation="KQTKLAFDMIAVAISAILTSHIRNADLNRCGPIIMWYFAR
    FISRMVFEVYRGNLIEEKTENYSIIPAIFLTVASFLENNFALSRRGAVFTLINF
    VLIVLFNVIKQFKDSPLFSTIYQKTLITTAERWENQVLFESHQIOKLVAVV
    LGTEIDKINLSPLYYVEAEIESTREVDHVFINLPSEFLDVQFVSDPELLIDV
    SVDINRGPFTALKKKKIQIDGHSITVESTNPKSHIMMKRLDILAVVGLICGI
    VSLVLPITRDGCPAIFRQKVGQNGRIFTTYKRSMTVVAEKKKOLLSONQMGV
    VCFKMGKLTLEQLDISYAKSLDELPQFVNLIDMSLVGTGRPTVDEPEKTLRQ
    KRRLSPKQITGLMQVSGRSNITDPDQVRLDLAYIDMWITMSDIKILKTVKVLRL
    EGSK"
gene
CDS
    /gene="cps2F"
    1374..1823
    /gene="cps2F"
    1374..1823
    /codon_start=1
    /transl_table=11
    /product="Cps2F"
    /protein_id="AAF18941.1"
    /db_xref="GI:6601340"
    /translation="MKVCLVSGSGGHLTHLYLKFPKKEERFWTTEDEKARSILKN
    EKMPCYFPTNRNLINLVNPTFLAFKILNDEKPDVITSSGAVAAPFFITGLFQAKT
    IYIEVDVRYNKSTLTKLVYPYTDIFIVQMEBKVKYPKSINLIGSIF"
    1823..2317
    /gene="cps1G"
    1823..2317
    /note="putative glycosyltransferase"
    /codon_start=1
    /transl_table=11
    /product="Cps1G"
    /protein_id="AAF18942.1"
    /db_xref="GI:6601341"
    /translation="MIFVTVGTHQGFNRLIKEIDLKNGSITDEIFIQTVSDYIP
    KYCKYKFLSYKMEQYVINKSEVVICGGPATFPMNLSKGGKQLPFRQKXGGEHND
    HOVEFRIRILQDNNILFIENIDLFKEKIEVSKQNFSTNNNPFCCRLKQIYKENEED
    QENE"
    3036..4202
    /gene="cps1H"
    3036..4202
    /note="putative capsular polysaccharide polymerase"
    /codon_start=1
    /transl_table=11
    /product="Cps1H"
    /protein_id="AAF18943.1"
    /db_xref="GI:6601342"
    /translation="MRKLFKYDEVFIPKYFMLIFPIPBOKYFLLIMNLIFHIKF
    LTKLILKNEILFLMSTICFYSVVTSPVEINFRRLPADPAPIMIAIMYVLY
    SPINIDYKDKNSIFPSFLVLGISALYIIQNKDVFDRHLIGLDYLTGYKTRIV
    GLMAYPFLNTTTTIVSIPLIFALIKNMQDFPLCLAFIPYISGSRISLSLAILII
    CLMRYIIGGFAMIKKLIIVFVILLIILNELLYHEILAVYNSRESNENARFIYOGS
    IDKVLNNILFGYGISSEVGTWLGSHSGYISFPYKSGIVGLILMFSPFYIKKSY
    GNVGTERALPFTSLAIFPIYETIDPIIIILVLFSSIGIMNNINPKDMETKNE"
    4195..5163
    /gene="cps1I"
    4195..5163
    /note="putative glycosyltransferase"
    /codon_start=1
    /transl_table=11
    /product="Cps1I"
    /protein_id="AAF18944.1"
    /db_xref="GI:6601343"
    /db_xref="GI:6601343"
```

```

/translation="MNDLSIVIPIVVVDYLDKDCINSIINQVTYNLEVLVNDSTP
DSEKICLNYAKNDRIKTYKKINGGLADANRGLGHEHNGKYLAFVSDSDYIEVAMPER
MNDNITEYNADIAIDFCLVDENGTYKGRNSNFVLTREBTYKEPESGSIENNNW
KLYSRDIIKQIKQINNRSIGEDLLEFLVNLNTRVAVDTRREYINIVIRNSSLINQ
KFSINNIDLVTRLNENYPPKLRERSHFADKVIKEKVKCLMKMSTQCLDNEFLPILF
SYRKEIRRYEPKAKRYLSRHLVTLYIMKFSPLLYMLYKFKQK"
5172..6143
/gene="cps1J"
/gene="cps1J"
5172..6143
/gene="cps1J"
/note="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cps1J"
/protein_id="AAF18945.1"
/db_xref="GI:6601344"
/translation="MDKISVLPVYVNDKYLSGCSIESIINQNYKRIEILLIDGSVD
SAKICKEAEKDRVKIPEFTHSGVSNARNNGIKRSLEYIMFVSDVDVSRILVEKL
YFNIIKRSNDSGLTYATFSENINNFEFNNDINDEPAINTVODMGEKFMPLYNNIF
STVCKLTKKRYITDLPQENOMLEDDLFNHLHYKINDRVSYLEHLYFVRGLISTV
NSFKGEVFLQLENLQKQVTVLTKQIYGEDFVSAIVKDTIRNQVYYISLMPKRYKQSI
FDKPLIFRNLKYKYFNLKVSNNKSNKFCIRIVSNKVKKILML"
6156..>6992
/gene="cps1K"
6156..>6992
/gene="cps1K"
/note="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cps1K"
/protein_id="AAF18946.1"
/db_xref="GI:6601345"
/translation="MDTISKISIIPIYVINYKYLSKCIDSIYNQTYKRIEILLVNDGS
TDNSEICLAIYAKKDSRIYRFRKNGGLSDARNGISAKQDYLAIFDSDFIHSFI
ORLHAIERENALVAVAGDYVDSGHFLTAEPLEPLTNQVAVSGRWCKLLEADGRIF
VVAQNKLYKKELFEDFRPEKGIHEDSEFTYRLYLEKVAIVKCLLYYVDRENSIT
TSSKTDHRFHCLBERQNRMDPYESRGKXELLBCYRFLAPVFLFKYNHMLSKQK
KKL"

ORIGIN
Query Match      100.0%; Score 6992; DB 1; Length 6992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGCGCAACGAATTTGGCATTATTTGATATAGACAGTTGCAATTTCTGCAATCTTAA 60
DB      1 ATGCGCAACGAATTTGGCATTATTTGATATAGACAGTTGCAATTTCTGCAATCTTAA 60
QY      61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAAATTTTATCATATAGATGCTTC 120
DB      61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAAATTTTATCATATAGATGCTTC 120
QY      121 ATTATTTTGCATTTTATATCTCGTATGCCAGTTGAATTTAGTATAGAGTAAATCTGA 180
DB      121 ATTATTTTGCATTTTATATCTCGTATGCCAGTTGAATTTAGTATAGAGTAAATCTGA 180
QY      181 TAGAGTTTGAAGAAAACATTAACTATAGTATATATTTTGCATTTTCTTACGGCAGTAT 240
DB      181 TAGAGTTTGAAGAAAACATTAACTATAGTATATATTTTGCATTTTCTTACGGCAGTAT 240
QY      241 CATTTTGTGGAGAAATAATTTGCGACTTTCAGAGCTGTGTCGCCGTATTTTCAACTTAA 300
DB      241 CATTTTGTGGAGAAATAATTTGCGACTTTCAGAGCTGTGTCGCCGTATTTTCAACTTAA 300
QY      301 TAAACTTGTGTTGGATATCTATTAAGTATATTTAAGCAGTTAAGATAGATGCTTC 360
DB      301 TAAACTTGTGTTGGATATCTATTAAGTATATTTAAGCAGTTAAGATAGATGCTTC 360
QY      361 TATTTTGCACATCTATCAAAAAAGACAGATTTCAATTAACAACGGCTGAACGATGGGAAA 420
DB      361 TATTTTGCACATCTATCAAAAAAGACAGATTTCAATTAACAACGGCTGAACGATGGGAAA 420
QY      421 ATATGCAAGTTTATTTGATATCATATAAACAAATTTCAAAAAAATCTTTGTCATTTGTTAG 480
```


Db	421	ATATCCAAAGTTTATTTGATCATAAACAATTCAAAAATCTGTGCATTGGTGG	480
Oy	481	TTTTAGGTACAGAAAATAGATAAAATTAATTATTCATACCGCTCAATTATCTGGGAAG	540
Db	481	TTTTAGGTACAGAAAATAGATAAAATTAATTATTCATACCGCTCAATTATCTGGGAAG	540
Oy	541	AAAGCTATATAGATTTCACACAGGGAAAGTGTGCACACGCTCTTATATAATCTACCAAGTG	600
Db	541	AAAGCTATATAGATTTCACACAGGGAAAGTGTGCACACGCTCTTATATAATCTACCAAGTG	600
Oy	601	AGTTTGTGACGTAAAGCAATTCGTTTCAGATTTGAGTTGATGTAGATATGATGTAAGCG	660
Db	601	AGTTTGTGACGTAAAGCAATTCGTTTCAGATTTGAGTTGATGTAGATATGATGTAAGCG	660
Oy	661	TTGATATATTAATTCATTCCGTTTCTACGCTGAAAAACAAAAAAATCCACAGCTAGGTC	720
Db	661	TTGATATATTAATTCATTCCGTTTCTACGCTGAAAAACAAAAAAATCCACAGCTAGGTC	720
Oy	721	ACCATAGCAATGTGAACTTTTTCACAAATTTTATTAAGCTTAGTCATATCATGATGAAC	780
Db	721	ACCATAGCAATGTGAACTTTTTCACAAATTTTATTAAGCTTAGTCATATCATGATGAAC	780
Oy	781	GACCTTTGATATCTCCGAGGGATAGTCGGGTAAATTTTGTGTAATAGTTTCTATTT	840
Db	781	GACCTTTGATATCTCCGAGGGATAGTCGGGTAAATTTTGTGTAATAGTTTCTATTT	840
Oy	841	TGTTAGTTCCAAATTAATCGTAGAGATGGTGACCGGCTATTTTCTCAGAAACGAGTTG	900
Db	841	TGTTAGTTCCAAATTAATCGTAGAGATGGTGACCGGCTATTTTCTCAGAAACGAGTTG	900
Oy	901	GACAGAAATGACGCAATATTTACATTCTACAAAGTTTCGATCGATGTATGTGATGCTGAGG	960
Db	901	GACAGAAATGACGCAATATTTACATTCTACAAAGTTTCGATCGATGTATGTGATGCTGAGG	960
Oy	961	AGCGCAAAAAAGACTTGTCTCAGCCCAAAACACAGATGCAAGGGTGGTATGTTTAAATGG	1020
Db	961	AGCGCAAAAAAGACTTGTCTCAGCCCAAAACACAGATGCAAGGGTGGTATGTTTAAATGG	1020
Oy	1021	GAAAAAGCATCCTACAAATTAATCCCAATTTGACATTTCATACGCAAAAAACAGTTTAGAG	1080
Db	1021	GAAAAAGCATCCTACAAATTAATCCCAATTTGACATTTCATACGCAAAAAACAGTTTAGAG	1080
Oy	1081	AGTTACCAACAGTTTATATATGTTTAAATTTGGCGATATGATCTAGTTGGTACACGCTCAC	1140
Db	1081	AGTTACCAACAGTTTATATATGTTTAAATTTGGCGATATGATCTAGTTGGTACACGCTCAC	1140
Oy	1141	CTACAGTTGATGAATTTGAAAAATATCTCTCGGTCAAAAGAGACGATTTAGTTTAAAC	1200
Db	1141	CTACAGTTGATGAATTTGAAAAATATCTCTCGGTCAAAAGAGACGATTTAGTTTAAAC	1200
Oy	1201	CAGGATTTACAGGCTCTCGGACAGGTTAGTGTCGTATGATATATCACACACTTCGACGACG	1260
Db	1201	CAGGATTTACAGGCTCTCGGACAGGTTAGTGTCGTATGATATATCACACACTTCGACGACG	1260
Oy	1261	TAGTTCCGTTGCACTTAGCATATCATTTGGAATTTGGAATTTGGAATTTGGAATTTGGAATTT	1320
Db	1261	TAGTTCCGTTGCACTTAGCATATCATTTGGAATTTGGAATTTGGAATTTGGAATTTGGAATTT	1320
Oy	1321	TATTAATAAGACAGTGAAGTTGTATTTGTTGAGAGAGGGAAGTAAGTAAAGTATATGAAG	1380
Db	1321	TATTAATAAGACAGTGAAGTTGTATTTGTTGAGAGAGGGAAGTAAGTAAAGTATATGAAG	1380
Oy	1381	TTTGTGTTGGTGGTCTTCACAGGGGACATTGACCTCTGTATTTGTTAAAAACGTTTT	1440
Db	1381	TTTGTGTTGGTGGTCTTCACAGGGGACATTGACCTCTGTATTTGTTAAAAACGTTTT	1440
Oy	1441	GGAAGCAAGAAAGACGTTTGTGGGTAAACATTTGATTAAGAGAGATCAAGAAAGTCTTTTGA	1500
Db	1441	GGAAGCAAGAAAGACGTTTGTGGGTAAACATTTGATTAAGAGAGATCAAGAAAGTCTTTTGA	1500
Oy	1501	AGAAATGAAAAATGTATCCATGTTACTTTCACAAATTCGCAATCTCATTAATTTAGTGA	1560

Db	1501	AGAAATGAAAAAATGATATCAATGTTACTTTCGAAACAAATCCGAATCTCATTTAATTAATGTA	1560
QY	1561	AAAATACTTTCTTAGCTTTCAAATTTTACGTATGAGAAACCAAGTTATTTATTCAT	1620
Db	1561	AAAATACTTTCTTAGCTTTCAAATTTTACGTATGAGAAACCAAGTTATTTATTCAT	1620
QY	1621	CTGGTGGCGGCGGTGCTGTCCCTCTTTTACATCGGAAAATCTATTTGGACCAAGACGA	1680
Db	1621	CTGGTGGCGGCGGTGCTGTCCCTCTTTTACATCGGAAAATCTATTTGGACCAAGACGA	1680
QY	1681	TTTATATGAGATATTTGATCGAGTAAATAATCTACATTAATCGGAAAAATCTAGTTTATC	1740
Db	1681	TTTATATGAGATATTTGATCGAGTAAATAATCTACATTAATCGGAAAAATCTAGTTTATC	1740
QY	1741	CCGTAAACAGATATTTTATTTGTTCAGTGGGAGAAATGAAGAAGTATATCTATAATCTA	1800
Db	1741	CCGTAAACAGATATTTTATTTGTTCAGTGGGAGAAATGAAGAAGTATATCTATAATCTA	1800
QY	1801	TTAATCTTGGGAGATTTTATTTATGATTTTGTAAACGTAGAGAACTCATGAACAAAGTT	1860
Db	1801	TTAATCTTGGGAGATTTTATTTATGATTTTGTAAACGTAGAGAACTCATGAACAAAGTT	1860
QY	1861	TAAATCGATGTGATAAAGAGATTGATTTTATTGAAAAAATGGAAGTATTAACCGACGAAT	1920
Db	1861	TAAATCGATGTGATAAAGAGATTGATTTTATTGAAAAAATGGAAGTATTAACCGACGAAT	1920
QY	1921	ATTTATTCMAACAGATATTTCTGAATATTTCCGAATATTTCCAGATATTAATAAATTTCT	1980
Db	1921	ATTTATTCMAACAGATATTTCTGAATATTTCCGAATATTTCCAGATATTAATAAATTTCT	1980
QY	1981	CAGTTACAAAGAAATGGAACATATATTTAAACAAATCAGAAAGTATTTTGGCCACGAGG	2040
Db	1981	CAGTTACAAAGAAATGGAACATATATTTAAACAAATCAGAAAGTATTTTGGCCACGAGG	2040
QY	2041	CCCCGCTACTTTTATGSAATTCATATATCCAAAGSAAAAACAATATTTGTTCCTAGCA	2100
Db	2041	CCCCGCTACTTTTATGSAATTCATATATCCAAAGSAAAAACAATATTTGTTCCTAGCA	2100
QY	2101	AAAAAGATGATGTAACATGTAAATGATCATCAAGTAGAGTTGTGTAAGAAATTTTACA	2160
Db	2101	AAAAAGATGATGTAACATGTAAATGATCATCAAGTAGAGTTGTGTAAGAAATTTTACA	2160
QY	2161	AGATPATPATATTTTATTTATAGAAAATATAGATGATTTGTTGAAAAAATATATGAACT	2220
Db	2161	AGATPATPATATTTTATTTATAGAAAATATAGATGATTTGTTGAAAAAATATATGAACT	2220
QY	2221	TTCTTAAGCAAACTAACCTTTACATCAAAATATATTTTTTTTGTGCAAGATTAAACAAT	2280
Db	2221	TTCTTAAGCAAACTAACCTTTACATCAAAATATATTTTTTTTGTGCAAGATTAAACAAT	2280
QY	2281	AGTTGAAGAAATTTTATGAGGATCAAGAAATGATATATTAATAAAGATGCAATTTGATTA	2340
Db	2281	AGTTGAAGAAATTTTATGAGGATCAAGAAATGATATATTAATAAAGATGCAATTTGATTA	2340
QY	2341	TGGCTTATCATTAATTTTCTCAGATTTTACTGAGAGGATACAGATATTTACATCTTCT	2400
Db	2341	TGGCTTATCATTAATTTTCTCAGATTTTACTGAGAGGATACAGATATTTACATCTTCT	2400
QY	2401	CTCAGAGAAATGCAACCAATTAAGTTCTCTCAGAAATCCTGTATATTAATTTTAAATATTC	2460
Db	2401	CTCAGAGAAATGCAACCAATTAAGTTCTCTCAGAAATCCTGTATATTAATTTTAAATATTC	2460
QY	2461	TCAGAGATTTATATGTGAATTTTACAAAAGATGAGCAAAAAATATTAAGAAAAATAGATATA	2520
Db	2461	TCAGAGATTTATATGTGAATTTTACAAAAGATGAGCAAAAAATATTAAGAAAAATAGATATA	2520
QY	2521	TGAACGAGTTAAATGTTACAGATATTTTCTTAATATATCAAGAAAAAATCTATGATAATGT	2580
Db	2521	TGAACGAGTTAAATGTTACAGATATTTTCTTAATATATCAAGAAAAAATCTATGATAATGT	2580
QY	2581	ACTGTTTAAAGATTTTATTAAGATGTATCGAGCTTTTGAATATCTATTTACAAAGATTGTT	2640
Db	2581	ACTGTTTAAAGATTTTATTAAGATGTATCGAGCTTTTGAATATCTATTTACAAAGATTGTT	2640

2641 GTTATTGATAGAAATAAAAACATGCTCTAAGAAATAGATTGGTCTTAATGGGTTG 2700
2641 GTTATTGATAGAAATAAAAACATGCTCTAAGAAATAGATTGGTCTTAATGGGTTG 2700
2701 CTTCACATGATTTTGTGGCAATCTTTTATCAAAATGAAAAAGCAACGCTTATTTAT 2760
2701 CTTCACATGATTTTGTGGCAATCTTTTATCAAAATGAAAAAGCAACGCTTATTTAT 2760
2761 AAGTATCTAAATGCGCAGATGAACCTATTATACAGCAATTAAGAAAAATGAAATT 2820
2761 AAGTATCTAAATGCGCAGATGAACCTATTATACAGCAATTAAGAAAAATGAAATT 2820
2821 TCAATATGATTAATCTAAATATGAAATTAAGATATATAAGTGAAGAAAAATCAACAT 2880
2821 TCAATATGATTAATCTAAATATGAAATTAAGATATATAAGTGAAGAAAAATCAACAT 2880
2881 TCTCTATTTGCTTTACAGATGATCTATTTGATGAAATGCTAAATGCAAGAAATTTAG 2940
2881 TCTCTATTTGCTTTACAGATGATCTATTTGATGAAATGCTAAATGCAAGAAATTTAG 2940
2941 TTTTATTTTGTGAGAAAGTTAAATAGAAATTAATCTAAATTTAAAGAAATTTTACT 3000
2941 TTTTATTTTGTGAGAAAGTTAAATAGAAATTAATCTAAATTTAAAGAAATTTTACT 3000
2941 TTTTATTTTGTGAGAAAGTTAAATAGAAATTAATCTAAATTTAAAGAAATTTTACT 3000
3001 AAAAAATTAATATGATTTTGTGAGATATGATGTTAAATTTAAATATATGACC 3060
3001 AAAAAATTAATATGATTTTGTGAGATATGATGTTAAATTTAAATATATGACC 3060
3001 AAAAAATTAATATGATTTTGTGAGATATGATGTTAAATTTAAATATATGACC 3060
3061 CGGAATATTTATTTTAAAGTCTTGTGATGATTTATTTTATTCAGAGCAAAAGTATG 3120
3061 CGGAATATTTATTTTAAAGTCTTGTGATGATTTATTTTATTCAGAGCAAAAGTATG 3120
3121 TATTTTATTTAATTTTATGAAATTTTATTTATTTATTTATTTTATTTTATTTTATTT 3180
3121 TATTTTATTTAATTTTATGAAATTTTATTTATTTATTTATTTTATTTTATTTTATTT 3180
3121 TATTTTATTTAATTTTATGAAATTTTATTTATTTATTTATTTTATTTTATTTTATTT 3180
3181 TAAATTTAAAAATGAAATTTTATGTTTATTTATTTATTTATTTATTTATTTATTTAT 3240
3181 TAAATTTAAAAATGAAATTTTATGTTTATTTATTTATTTATTTATTTATTTATTTAT 3240
3181 TAAATTTAAAAATGAAATTTTATGTTTATTTATTTATTTATTTATTTATTTATTTAT 3240
3241 TAGTCACAGATGTTTGTGAAATTAATTTTGAAGATTTATTTGCAATTTTACTGCTC 3300
3241 TAGTCACAGATGTTTGTGAAATTAATTTTGAAGATTTATTTGCAATTTTACTGCTC 3300
3301 CCATATTTTGAATTTATGCAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT 3360
3301 CCATATTTTGAATTTATGCAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT 3360
3301 CCATATTTTGAATTTATGCAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT 3360
3361 AAAAAATTTAAAAATAGATCTTTTATTTATTTTATTTATTTATTTATTTATTTATTTAT 3420
3361 AAAAAATTTAAAAATAGATCTTTTATTTATTTTATTTATTTATTTATTTATTTATTTAT 3420
3421 TGTATTTATTTCAAAATGGAAGATTTATTTTATTTATTTATTTATTTATTTATTTATTT 3480
3421 TGTATTTATTTCAAAATGGAAGATTTATTTTATTTATTTATTTATTTATTTATTTATTT 3480
3481 ACTATCTTAAACAGGCGTCAAAACAAAGTGTGCTTATGAACTATTCCTAACCTTAA 3540
3481 ACTATCTTAAACAGGCGTCAAAACAAAGTGTGCTTATGAACTATTCCTAACCTTAA 3540
3541 ATACCACTTCAAAATATGTTTCAATTCGTTATCTTTGACCTTATTAATAAATTAATATG 3600
3541 ATACCACTTCAAAATATGTTTCAATTCGTTATCTTTGACCTTATTAATAAATTAATATG 3600
3601 AACCAATTTTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGT 3660
3601 AACCAATTTTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGT 3660
3661 GTAGTTTATCGTAGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3720
3661 GTAGTTTATCGTAGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3720
3661 GTAGTTTATCGTAGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3720

3721 TTGCTTGGATTAATAAAGCTATAGTATATTTGTATATACACTTATTTATTTATTTATCTG 3780
3721 TTGCTTGGATTAATAAAGCTATAGTATATTTGTATATACACTTATTTATTTATTTATCTG 3780
3781 AATGCTTTACATGAATTTTGTGCTTTTATTAATTTCTAGAGATCAAGTAAACGAAGCTA 3840
3781 AATGCTTTACATGAATTTTGTGCTTTTATTAATTTCTAGAGATCAAGTAAACGAAGCTA 3840
3841 GATTATTTATTTATCAAGAAATTTATTAATTTATTTATTTATTTATTTATTTATTTATTT 3900
3841 GATTATTTATTTATCAAGAAATTTATTAATTTATTTATTTATTTATTTATTTATTTATTT 3900
3841 GATTATTTATTTATCAAGAAATTTATTAATTTATTTATTTATTTATTTATTTATTTATTT 3900
3901 ATGGAATTCGGAATTTATTCAGATTCGGAATTCGGAATTCGGAATTCGGAATTCGGAAT 3960
3901 ATGGAATTCGGAATTTATTCAGATTCGGAATTCGGAATTCGGAATTCGGAATTCGGAAT 3960
3901 ATGGAATTCGGAATTTATTCAGATTCGGAATTCGGAATTCGGAATTCGGAATTCGGAAT 3960
3961 CATTTTATTTATTAATTCAGAAATTTAGTGGTGAATTTTATTTATTTATTTATTTATTTAT 4020
3961 CATTTTATTTATTAATTCAGAAATTTAGTGGTGAATTTTATTTATTTATTTATTTATTTAT 4020
4021 TTAATTAATAAAGTATGAGATTAATGGGAAACAGACTATTTTATTTATTTATTTATTTAT 4080
4021 TTAATTAATAAAGTATGAGATTAATGGGAAACAGACTATTTTATTTATTTATTTATTTAT 4080
4081 CCATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4140
4081 CCATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4140
4141 CTTCAATGATTTTGAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4200
4141 CTTCAATGATTTTGAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4200
4201 GATTTAATTTATGATTTATGCAATTTATTAATTTATTTATTTATTTATTTATTTATTTAT 4260
4201 GATTTAATTTATGATTTATGCAATTTATTAATTTATTTATTTATTTATTTATTTATTTAT 4260
4201 GATTTAATTTATGATTTATGCAATTTATTAATTTATTTATTTATTTATTTATTTATTTAT 4260
4261 AACGATTTATTTATCAAAACATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 4320
4261 AACGATTTATTTATCAAAACATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 4320
4321 ACTGATGATTTGAGAAATTTGCTTAACTATTAAGAAAGATGAAAGATTTAAATAT 4380
4321 ACTGATGATTTGAGAAATTTGCTTAACTATTAAGAAAGATGAAAGATTTAAATAT 4380
4321 ACTGATGATTTGAGAAATTTGCTTAACTATTAAGAAAGATGAAAGATTTAAATAT 4380
4381 TACAGAAATTTAATGGCGCTAGCAGATGCTGAAATTTGCGACTAGAAACATGCAACA 4440
4381 TACAGAAATTTAATGGCGCTAGCAGATGCTGAAATTTGCGACTAGAAACATGCAACA 4440
4441 GGTAAATATTTGCTTTTGTGATTTCTGATGCTATTAAGAAAGATTTGCTTAACT 4500
4441 GGTAAATATTTGCTTTTGTGATTTCTGATGCTATTAAGAAAGATTTGCTTAACT 4500
4441 GGTAAATATTTGCTTTTGTGATTTCTGATGCTATTAAGAAAGATTTGCTTAACT 4500
4501 ATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4560
4501 ATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4560
4501 ATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4560
4561 GACGAAACGCGGTATTAAGAAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4620
4561 GACGAAACGCGGTATTAAGAAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4620
4621 GAGACTGTAAAGAAATTTTGTCAAGATCTAATTAAGAAATTTATTTATTTATTTATTTAT 4680
4621 GAGACTGTAAAGAAATTTTGTCAAGATCTAATTAAGAAATTTATTTATTTATTTATTTAT 4680
4681 TATTCACGAGATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4740
4681 TATTCACGAGATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4740
4681 TATTCACGAGATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4740
4741 GATTTGCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4800
4741 GATTTGCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4800
4741 GATTTGCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4800
4801 GATATTTATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4860

```
Db 4801 GAATATATTATATATATGATCATCGTAACAGTTCGCTATATATCAAGAAATTCCTATA 4860
Qy 4861 AATAATATGATTTAGACAGAGATGAGAAATTAACCCCTTAAGTAAAAAGAGATT 4920
Db 4861 AATAATATGATTTAGACAGAGATGAGAAATTAACCCCTTAAGTAAAAAGAGATT 4920
Qy 4921 AGCATATATTTTGAAGCAAGATTTAAAGAGAGTTAAATGTTTAAACCAATGTAT 4980
Db 4921 AGCATATATTTTGAAGCAAGATTTAAAGAGAGTTAAATGTTTAAACCAATGTAT 4980
Qy 4981 TCAACAGATTTGTTGATTAATGAGTTCCTGCAATATTAAGTCTTAACGAAAAGATA 5040
Db 4981 TCAACAGATTTGTTGATTAATGAGTTCCTGCAATATTAAGTCTTAACGAAAAGATA 5040
Qy 5041 CGTAGATATCCATTATTAAGCGAAAAGATATTTATCAAGAAAGCATTTAGTACGTTG 5100
Db 5041 CGTAGATATCCATTATTAAGCGAAAAGATATTTATCAAGAAAGCATTTAGTACGTTG 5100
Qy 5101 TATTGATGAATTTTGCCCTAACTATATGTAATGTTATTAAGAAATTTCAAGACAG 5160
Db 5101 TATTGATGAATTTTGCCCTAACTATATGTAATGTTATTAAGAAATTTCAAGACAG 5160
Qy 5161 TAGAGTAAATATGATAAATTAAGTATATGTTCCAGTTTATATATGATAAATAT 5220
Db 5161 TAGAGTAAATATGATAAATTAAGTATATGTTCCAGTTTATATATGATAAATAT 5220
Qy 5221 TAAAGTATGATTAAGAAAGCATTAATCAAAAATTAATAAATATGAAATATTTGA 5280
Db 5221 TAAAGTATGATTAAGAAAGCATTAATCAAAAATTAATAAATATGAAATATTTGA 5280
Qy 5281 TAGATGATGCTGCTGATGATGATTCGCTAAATATGCAAGAAATATGCAAAAAGATA 5340
Db 5281 TAGATGATGCTGCTGATGATGATTCGCTAAATATGCAAGAAATATGCAAAAAGATA 5340
Qy 5341 AAAGAGTAAATATTTTTCATCATATAGTATGAGATCAAAATGATCAAGAAATCATGAA 5400
Db 5341 AAAGAGTAAATATTTTTCATCATATAGTATGAGATCAAAATGATCAAGAAATCATGAA 5400
Qy 5401 TAAAGCGAGTACAGCTGAATATATATGTTGTTGACCTGATGATGTTGTTAGTAA 5460
Db 5401 TAAAGCGAGTACAGCTGAATATATATGTTGTTGACCTGATGATGTTGTTAGTAA 5460
Qy 5461 GATTAGTAAAGAAATTAATTTTATATTAATTAATAAAGTAAAGTATTAATGCGTGT 5520
Db 5461 GATTAGTAAAGAAATTAATTTTATATTAATTAATAAAGTAAAGTATTAATGCGTGT 5520
Qy 5521 TGTACGCTACTTTTTCAGAAATATTAATATTTTGAAGTGAATATCCAAATTTGAT 5580
Db 5521 TGTACGCTACTTTTTCAGAAATATTAATATTTTGAAGTGAATATCCAAATTTGAT 5580
Qy 5581 TTGAAGCAATTAATACCGTCAAGACATGGAGAAAAAATTTTATGAAATTTGTATTA 5640
Db 5581 TTGAAGCAATTAATATACCGTCAAGACATGGAGAAAAAATTTTATGAAATTTGTATTA 5640
Qy 5641 ATATATATTTTTCATCTCTGTTTGTAACTATATTAAGAAAGTAAATCAATACGATCTT 5700
Db 5641 ATATATATTTTTCATCTCTGTTTGTAACTATATTAAGAAAGTAAATCAATACGATCTT 5700
Qy 5701 TTCAAGAGATCAATGTTAGAGAAAGATTTTATCTTTTATCTGATTTTAAAGATA 5760
Db 5701 TTCAAGAGATCAATGTTAGAGAAAGATTTTATCTTTTATCTGATTTTAAAGATA 5760
Qy 5761 TAGATAGAGTATGATTTAGTACATGAAATCTTTATTTTATAGAGAGTATCTAAGTA 5820
Db 5761 TAGATAGAGTATGATTTAGTACATGAAATCTTTATTTTATAGAGAGTATCTAAGTA 5820
Qy 5821 CAGTAAATCTTTTAAAGAGTGTGTTTGAATTTGCAAAATTTGCAAAATCAATGA 5880
Db 5821 CAGTAAATCTTTTAAAGAGTGTGTTTGAATTTGCAAAATTTGCAAAATCAATGA 5880
Qy 5881 TAGATATGTTTAAAGCAATATATGATGAGATTTTGAAGTATCAATGTTTAAAGATACTA 5940
Db 5881 TAGATATGTTTAAAGCAATATATGATGAGATTTTGAAGTATCAATGTTTAAAGATACTA 5940

Db 5881 TAGATATGTTTAAAGCAATATATGATGAGATTTTGAAGTATCAATGTTTAAAGATACTA 5940
Qy 5941 TAGCTGGCAAGTATTTTATATAGCTTAATATGTTTAAATACGAAAACAGCTATTT 6000
Db 5941 TAGCTGGCAAGTATTTTATATAGCTTAATATGTTTAAATACGAAAACAGCTATTT 6000
Qy 6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTATTAATAAATATTAATTTTAACTGTTAAAG 6060
Db 6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTATTAATAAATATTAATTTTAACTGTTAAAG 6060
Qy 6061 TATCTAACAAAATTTCTTGTCTTAATAATTTTGTATTAAGATTTGTTCCAAAGATT 6120
Db 6061 TATCTAACAAAATTTCTTGTCTTAATAATTTTGTATTAAGATTTGTTCCAAAGATT 6120
Qy 6121 TTTAAAAAATTAATATGTTTAAATAGGAGATATCAATGATCATATTAAGTAAATTTCTA 6180
Db 6121 TTTAAAAAATTAATATGTTTAAATAGGAGATATCAATGATCATATTAAGTAAATTTCTA 6180
Qy 6181 TAAATGACCTATATATATATGTAAGAAATATTTATCTAATGATATAGATTAAGATTGTA 6240
Db 6181 TAAATGACCTATATATATATGTAAGAAATATTTATCTAATGATATAGATTAAGATTGTA 6240
Qy 6241 ATGAGACTTACAAACATATAGAGATTTCTGCTGAATGACGTTAGTCGATTAATTCG 6300
Db 6241 ATGAGACTTACAAACATATAGAGATTTCTGCTGAATGACGTTAGTCGATTAATTCG 6300
Qy 6301 AAGAAATTTGTTAGCATATGCGAAGAAATATGTCGATTCGTTATTTTAAAAAAGATA 6360
Db 6301 AAGAAATTTGTTAGCATATGCGAAGAAATATGTCGATTCGTTATTTTAAAAAAGATA 6360
Qy 6361 ACGCGGGCTATGAGATGCGCGTAAATTAATGAGCAATATGCGCGCAAGGTTGACTATAG 6420
Db 6361 ACGCGGGCTATGAGATGCGCGTAAATTAATGAGCAATATGCGCGCGCAAGGTTGACTATAG 6420
Qy 6421 CTTTATATAGCTAGATGATTTTATTCATTCGAGATTCATCAACGTTTACAGAACAA 6480
Db 6421 CTTTATATAGCTAGATGATTTTATTCATTCGAGATTCATCAACGTTTACAGAACAA 6480
Qy 6481 TTGAGAGAGAGATGCGCTTGTGTGAGTGTCTGTTATGATAGGTATGATGCTTGGGGC 6540
Db 6481 TTGAGAGAGAGATGCGCTTGTGTGAGTGTCTGTTATGATAGGTATGATGCTTGGGGC 6540
Qy 6541 ATTTCTTAAACAGAGAGCGCTCTCTAACAATCAGCGCTGTTGAGCGGACAGAAATGTT 6600
Db 6541 ATTTCTTAAACAGAGAGCGCTCTCTAACAATCAGCGCTGTTGAGCGGACAGAAATGTT 6600
Qy 6601 GTAAAAAGCTGCTAGAGCGGATGCTCATGCTTGTGTGCGCTGTAAATPAACTCTATA 6660
Db 6601 GTAAAAAGCTGCTAGAGCGGATGCTCATGCTTGTGTGCGCTGTAAATPAACTCTATA 6660
Qy 6661 AAAAAAGACTATTTGAAGATTTTCGATTTGAAAAAGGTAAAGATTCATGAAGATGAATACT 6720
Db 6661 AAAAAAGACTATTTGAAGATTTTCGATTTGAAAAAGGTAAAGATTCATGAAGATGAATACT 6720
Qy 6721 TCACTTATGCTGCTCTATAGATTTGAAAAAGTTCGAATAGTTAAGAGATGCTGTACT 6780
Db 6721 TCACTTATGCTGCTCTATAGATTTGAAAAAGTTCGAATAGTTAAGAGATGCTGTACT 6780
Qy 6781 ATTATGTTGACCGAGAAATATGATACAACTTCTAGCATGATGATCCATGCTTCATT 6840
Db 6781 ATTATGTTGACCGAGAAATATGATACAACTTCTAGCATGATGATCCATGCTTCATT 6840
Qy 6841 GCTTACTGGAATTTCAAAATGAACGAATGACCTTCTATGAAGATTAAGAGATTAAGAGC 6900
Db 6841 GCTTACTGGAATTTCAAAATGAACGAATGACCTTCTATGAAGATTAAGAGATTAAGAGC 6900
Qy 6901 TCTTACTAGAGTGTATGCTTCAATTTTGTAGCCTTGTGCTTTGTTTGGCAAAATATA 6960
Db 6901 TCTTACTAGAGTGTATGCTTCAATTTTGTAGCCTTGTGCTTTGTTTGGCAAAATATA 6960
Qy 6961 ATCATGTTGAGCAAAACAGCAAAAGAGCTT 6992
Db 6961 ATCATGTTGAGCAAAACAGCAAAAGAGCTT 6992
```

RESULT 2	LOCUS	DEFINITION	ACCESSION	KEYWORDS	SOURCE	REFERENCE	TITLE	JOURNAL	FEATURES
AF118389	15401 bp	DNA linear BCT 11-Apr-1999							
AF118389		Streptococcus suis Cps2A (cps2A), Cps2B (cps2B), Cps2C (cps2C), Cps2D (cps2D), Cps2E (cps2E), Cps2F (cps2F), Cps2G (cps2G), Cps2H (cps2H), Cps2I (cps2I), and Cps2J (cps2J) genes, complete cds; Cps2K (cps2K) gene, partial cds; and unknown genes.							
AF118389									
AF118389.1	GI:4580620								
		Streptococcus suis							
		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.							
	1 (bases 1 to 15401)								
	Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.O., Stockhofe-Zurwieden,N. and Smits,M.A.								
	Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor								
	Infect. Immun. 67 (4), 1750-1756 (1999)								
	10085014								
	2 (bases 1 to 15401)								
	Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.O., Stockhofe-Zurwieden,N. and Smits,M.A.								
	Direct Submission								
	Submitted (07-JAN-1999) Department of Bacteriology, DLO-Institute for Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The Netherlands								
	Location/Qualifiers								
	1..15401								
	/organism="Streptococcus suis"								
	/mol_type="genomic DNA"								
	/strain="10"								
	/serotype="2"								
	/db_xref="taxon:1307"								
	<1..721								
	/note="Ori22"								
	/codon_start=2								
	/transl_table=11								
	/product="unknown"								
	/protein_id="AAD24444.1"								
	/db_xref="GI:4580621"								
	/translation="SLDIDHMEVMEASKASAGSACSPQAYQAFFGAENIYVTTGGLSGSFNAARVARDMYIEEPNNVTHLIDLSASGEMDLVHOINLISAGDFPOVVEALITHREHSLFLPVLAKEVNLVNGRSLVGVVGLINIRVGAASAEKLELLQVARGHKGVTAAFEEMKAGYDGGRIVMAHNNKAFQGFSELVAKSFPAVIDEVAISGLGSFYAEGGLMGVEVKA"								
	complement (820..2079)								
	/note="Ori21"								
	/codon_start=1								
	/transl_table=11								
	/product="unknown"								
	/protein_id="AAD24445.1"								
	/db_xref="GI:4580622"								
	/translation="MKKQVITIQDILTIEEHRFRGKLPSTIRLAEOVHCSMDTVQKAMLELKQNKIYAVEKSGYILIEDRFDHTRCRQSTRLSRITTEPRICKSLGLGRNYLFNNYHQOEGLAELISSVQSLMDYHYVTKRQQLVITAGSQALYILTOMETLAKTELLINENPTYSRMIELIRHQGIPIYQTIERNLDGIDEBESIFQCGKIKFYTIIPRLHNPGSTYDIAITKTAIVLAKQYDVYIEEDYADPSSHSLPLHYLDENRVIYIKSPTPLPAPARIGALISPLNOADIFIKKSLIDVPTNLMOKALSYLDNEMFARNQGHHTHYAOMNKIKDCEKYALNIPYRIPKSGVTFQSLSKTISLPSIQHMFQKCYVFSGGQKADFLQTFEODFPADKLEQFVRYLNE"								
	2202..2936								
	/note="Ori2X"								
	/codon_start=1								
	/transl_table=11								
	/product="unknown"								
	/protein_id="AAD24446.1"								
	/db_xref="GI:4580623"								
	/translation="MKLIIPNNKEVNTLENNASYLLSDRSKPVLDALISQDFVKKMAA								

[illegible]

```
/codon_start=1
/transl_table=11
/product="Cpe22"
/protein_id="AAD24451.1"
/db_xref="GI:4580628"
/transl_table=11
/transl_table=11
/translation="MNIETGYRQTKLALPDMIAVTISALITSHIPNADINRSGITFIIM
MNYAFEFISRMVEFEYRGNLIEFKETVYSIIPIVILMAVSEPMLENNPALSRRAV
YPTLINFVLYVLEFVLEIKQKOSFLSTYOKKIIITLALMENNQVLESDIIEFOK
NIVALYITGEIDKINPLPLYVSEAEAGFSTREVDVYFVILPSEYDLOKVSDF
ELIGDIVGDINSRGFTVLKNKKIOMLGHSIYTBSTNYPKSHIMKKLIDIGAV
GLISGIVSILILPIITRROGPAIPAKRVGONGRIFTYKRSFVDAEVRKEILMA
QUNQMGFMKMDPRITPIGHFIRKTSIDELPQFYNAVIGDMSIVGTPTVDEDEK
YTPSQRRRLSFKFGITGLTWQSGRSDITDFNEVRLDITVIMSDIKILKLTVK
VVLREGQ"
/gene="Cpe22"
/gene="Cpe22F"
/8089..9258
/gene="Cpe22F"
/8089..9258
/feature="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cpe22F"
/protein_id="AAD24452.1"
/db_xref="GI:4580629"
/translation="MRTVYIGSGKIPAKYGFETFEVKLTETXOKDSINVFACITRE
NSAKSDITGEVFEHNGATCPNIDVPNGSKAILYDIMALKKSIEIAKDRDTPSIFY
IACRIGPRTYLEFKQIBESTIGGLPVNPGHEMLRKMVSYPRAQYKSESLMKYAD
LILCDSKNIETKIHEDEYKRYAPETSIATGTDLDKSLSPDSVAREWKKEISND
YVLVGRFVPENNYEWIREFMKSYSRKQFVLTINVEHNSFYELKKEGFPDKKIK
FVGVVNOELKXIRENAFAYFHGEVGGTNPISLEALSTKLMLLDVGFREYVEE
GAKYMKDMLHRIYDSCQEQLSQEQINDMSLSTKQYKERFSMDFIVDEYEKLEPKG"
/gene="Cpe22"
/gene="Cpe22"
/9262..10419
/gene="Cpe22"
/9262..10419
/feature="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cpe22"
/protein_id="AAD24453.1"
/db_xref="GI:4580630"
/translation="MKKILYLHAGAEIYAGDKVLELILKGLDKNEFEAHVILPNDGL
VVALREGVQVEVINYPILRKYFNPKGIDFYFISYHYSKOIAYAIENKVDIHHN
TTAVLEGIVLKRKLKILMHVHEITVKKFTISDSINPLMGRPADKIYVSOAVANHI
KQSPHDKDQISVITNGVDNRYFGSDASVAREPIDEALYIGMVGKVMKQGD
FLAAVAPILQNPKALFIAGSAFBESENRVLEKISIQLVSSQVRMDVYANTE
LYNMDIFVLPTNPDPPTVYLKAMACGKPVYGVHGVCEMVEGKVGVPITPNSP
LMSKVIQLDSINILRRKIGNNSIEROKEHFSLSKYVNFVKVYTSIKY"
/gene="Cpe22"
/gene="Cpe22H"
/10808..12178
/gene="Cpe22H"
/10808..12178
/feature="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cpe22H"
/protein_id="AAD24454.1"
/db_xref="GI:4580631"
Query Match 8.7%; Score 611; DB 1; Length 15401;
Best Local Similarity 9.6%; Pred. No. 5.8e-247;
Matches 761; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Db 14758 TTTAAAAAGAGAACGGCGGGCTATCAGATGCCCGTAATTATGACATTAAGTCGGCCCAAG 14817
Qy 6409 GTCAGTACTAGCTTTTATAGACTGATGATTTTATTCATTCGAGTTTCATCAACGTT 6468
Db 14818 GTGACTACTTAGCTTTTATAGACTGATGATTTTATTCATTCGAGTTTCATCAACGTT 14877
Qy 6469 TACACAGAACATTTGAGAGAGAAATGCCCTTGCGAGTGTGCTTATGATAGAGGTAG 6528
Db 14878 TACAGAACCAATTGAGAGAGAAATGCCCTTGCGAGTGTGCTTATGATAGAGGTAG 14937
Qy 6529 ATGCTTGGGGCATTTTCTTAACGACGAGCCCTTCTCAAAATCGAGCTGTTCGAGCG 6588
Db 14938 ATGCTTGGGGCATTTTCTTAACGACGAGCCCTTCTCAAAATCGAGCTGTTCGAGCG 14997
Qy 6589 GACGAAATGTTGTAAAGAGCTGCTAGAGCGGATGCTATGCTTGTGTGCGCTGTA 6648
Db 14998 GACGAAATGTTGTAAAGAGCTGCTAGAGCGGATGCTATGCTTGTGTGCGCTGTA 15057
Qy 6649 ATAACTCTATTAAGAGAACTATTTGAAGATTTTGAATTTGAAAAGGTAAGATTCAAG 6708
Db 15058 ATAACTCTATTAAGAGAACTATTTGAAGATTTTGAATTTGAAAAGGTAAGATTCAAG 15117
Qy 6709 AAGATGAATATCTTCACTATTCGCTTCTCTATGAGTTAGAAAAAGTTGCAATGTTAAG 6768
Db 15118 AAGATGAATATCTTCACTATTCGCTTCTCTATGAGTTAGAAAAAGTTGCAATGTTAAG 15177
Qy 6769 AGTCTTGTACTATTAATGTTGACCGAGAAATAGTATCAACTTCTAGCATGACGACC 6828
Db 15178 AGTCTTGTACTATTAATGTTGACCGAGAAATAGTATCAACTTCTAGCATGACGACC 15237
Qy 6829 ATGCTTGTACTATTAATGTTGACCGAGAAATAGTATCAACTTCTAGCATGACGACC 6888
Db 15238 ATGCTTGTACTATTAATGTTGACCGAGAAATAGTATCAACTTCTAGCATGACGACC 15297
Qy 6889 GAGATTAAGAGCTCTTACTAGAGTGTATCGTTCAATTTTATGAGCTTGTGTTGTTT 6948
Db 15298 GAGATTAAGAGCTCTTACTAGAGTGTATCGTTCAATTTTATGAGCTTGTGTTGTTT 15357
Qy 6949 TAGGCAATATTAATCTTGTGTTGAGCAACAGCAAAAGAAAGCTT 6992
Db 15358 TAGGCAATATTAATCTTGTGTTGAGCAACAGCAAAAGAAAGCTT 15401
RESULT 3
LOCUS AY057915 17468 bp DNA linear BCT 01-NOV-2002
DEFINITION Streptococcus thermophilus eps type III operon, partial sequence.
ACCESSION AY057915
VERSION AY057915.1 GI:24473733
KEYWORDS Streptococcus thermophilus
SOURCE Streptococcus thermophilus
ORGANISM Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 17468)
AUTHORS Rallu,F., Ehrlich,D.S. and Renault,P.
TITLE Diversity of eps operons in Streptococcus thermophilus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 17468)
AUTHORS Rallu,F., Ehrlich,D.S. and Renault,P.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Genetique Microbiome, INRA, Domaine de Vilvert, Jouy en Josas 78352, France
FEATURES
source
1..17468
/organism="Streptococcus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:1308"
<1..17075
/gene="eps type III operon"
<1..492
/gene="deop"
<1..492
CDS
```

/gene="deod"
/codon_start=1
/transl_table=11
/product="deod"
/protein_id="AAL23725.1"
/db_xref="GI:24473734"
/translation="RELIYDYGKKLIRVGTAGSLNEDVHRELVLAAQAAATNSNIIIR
NDMPDYDFPOLASFNLIDKAVHIAKNFGMTTHGVLSDFVSYNFEKNI ELGKMGV
KAVEMEAALYYLAAQHVDALAIWITI SDLSLVNPDDEPTAERQNTFTDMKVGLETL
IAD". .2288
/gene="eps3a"
828. .2288
/gene="eps3a"
/codon_start=1
/transl_table=11
/product="Eps3a"
/protein_id="AAL23726.1"
/db_xref="GI:24473735"
/translation="MSSRTNRKQKRTGNRSGMGNVGLTILYAILALVLLFTMFNYP
LSFRPLNITITIGLVLAISIFLOKTKSPVTYVVLVIFSLVSLVGI FGKQIMDI
TMRMGTAFSEBEMSIYVPKESDIKDVSOLTSVOAPTKVDKNITEILMSALCKKIV
DYKVDVASIYQAIYDNLKSGSKAWLSGSYASLESVDNFPASNKXTLYTKIKKKV
NNSANQVDSKVFNIYISGIDTYGSISTVSRSDVNIITVMNMTHKILLITTRDXYK
IPGGANQYDKLTHAGIYGVETSEQTLDELYGTKIDYARINFTSLKILIDLGATV
HNDQAFSLHGKDFPVGDIQWNSBOALFVBERYSIDGDNDRGNQBRKVSALVVK
LASLKSNSFTSIVNNLQDSVOTNISLDTINLANLTQDSGSKFTVTSQAVGTGSGTG
QLTSYAMPNSSLYMKLQNSVASASQAIKNIMEEK"
2289. .3020
/gene="eps3b"
2289. .3020
/gene="eps3b"
/codon_start=1
/transl_table=11
/product="Eps3b"
/protein_id="AAL23727.1"
/db_xref="GI:24473736"
/translation="MIDVHSHIVFDVDDGPKTLESLDLIGESVAGVKKIVSTSHRR
KCMFEPBEDKIFANFSKVAEALVPDLITYGGLYYTLDI VEKLEKNLLPRMNT
OPALIRFSA RTSWKETHSGLSVLRAGVPIVATHRYDALNEADRVREIINMGCT
QVNSHVLRKLFGDGKEKVRKKRVRPFLEKNLVHVMVASDMHNLGPRPPMKQAYELVK
KNYGSKRANLFIENPKTLENLQYL"
3029. .3721
/gene="eps3c"
3029. .3721
/gene="eps3c"
/codon_start=1
/transl_table=11
/product="Eps3c"
/protein_id="AAL23728.1"
/db_xref="GI:24473737"
/translation="MNQDNTKSDEIIVLALLHKLMTKULLLFTAFYPAASFLLGYF
FIQPIYTSIRITVYNQATDNNLSKQDLOAGTYLLANDYKEITTSNDVLSVYKDEKL
NISEALSKWVSNIPTDTRLISIVNAKQQAQTLANKVRVASAKKIKVYKVEDV
TMLBEAKLPESPSPSNIKLVNLGAVLGFLAVGVLVREIILDDRVRREDVDALGM
ALLGIVPDYDKI"
3731. .4489
/gene="eps3d"
3731. .4489
/gene="eps3d"
/codon_start=1
/transl_table=11
/product="Eps3d"
/protein_id="AAL23729.1"
/db_xref="GI:24473738"
/translation="MPLKLVKSRVNFQKTEEYVNAIRTIQFSGAQIKVIAISSVE
AGEGKSTISLNLAISFASVGLRTLIDADRNSVFSCTFKNSEPYKGLISNPLSGNDL
NETICOTDISGLDVIASGVPNPTSLQDNDRNPLHMEVARSRYDIIDTPPGVAVI
DAYTAAQADASILLTPACKIKRPFYTKAIEQCELOSSQGLGVVANKVDMTVKKYSY
GSYGSYGSYSYGEYRKKTDTQTEGSHAAHRRRG"

/gene="eps3e"
/note="undecaprenylphosphate glucosyltransferase"
/codon_start=1
/transl_table=11
/product="Eps3e"
/protein_id="AAL23730.1"
/db_xref="GI:24473739"
/translation="MKQDEIRRIEIGIQLVWVIFPAMVASKI PYTEITQGSITVLG
VWVHSYIISYENLKYRGYLDLILATKCFIPALATLSPFPADSPSISRGLL
VYVMSIGVLVYNTVLKYSRSLYTRKRSKNLILLSDQARLDNLSRMKQNDGRI
SAVCVLDNPPYPTDPEFIKSKDPENLIEVATHSVNDQVILNLPSEQIKIMDVASPELWV
IPVGINNALPEFSGEKHIQOLGPFKVTSTHFYSYDILLAKRFLDICALLVGLV
CGIYCIPLYLIRKDGGAIFAORYVGENGRIFPKYFRSRKRVDAEELKKMLMONOM
SGMFKMDNPRITKIGHFIKRTSLDELPQFWNVLYKGMSLVGTFRPPTVDEYKXTP
QKRRLSFRKPGITGLMVSGRSEITPDEVVLQVAYIDGWTIMRDIQILLTKIVVM
KDGAK"
5946. .6264
/gene="orf3a"
/pseudo
5946. .6264
/gene="orf3a"
/note="orf3a"
/pseudo
/codon_start=1
/transl_table=11
/gene="eps3f"
/pseudo
6275. .7647
/gene="eps3f"
/note="eps3f"
/note="eps3f, putative undecaprenylphosphate
glycosyltransferase"
/pseudo
/codon_start=1
/transl_table=11
/gene="eps3g"
/pseudo
7682. .8131
/gene="eps3g"
7682. .8131
/gene="eps3g"
/note="putative galactosyltransferase enhancer"
/codon_start=1
/transl_table=11
/product="Eps3g"
/protein_id="AAL23731.1"
/db_xref="GI:24473740"
/translation="MKICVSGSGHLAHLNMLKPFMKDEERFWTFDKEDARSILKD
EKMYGCHPEPTNRNLKNLKNPFLAIKVLKEKEDVLISSGAVAVFPFYGKLPFAKT
VYIEVFDRI DKPTVYGLVYPVTDKFIYQWEMKQVYPRALVLSIF"
8131. .8637
/gene="eps3h"
8131. .8637
/gene="eps3h"
/note="putative galactosyltransferase"
/codon_start=1
/transl_table=11
/product="eps3h"
/protein_id="AAL23732.1"
/db_xref="GI:24473741"
/translation="MIFVYGTHEQEPFNRLIOEVHLVETGVIKEEVFIQTGYSIYEP
KTCQMSRLISPQMSKPFQKADIIITHGSPATFMSAITNGKKPIVVPQEKGEHND
HOVDPRANVAKRMGTIILVEDVDKLGFTIENINSIVALEIYMGNNKKFNNDDLEGI
VRLDNVX"
8642. .9358
/gene="eps3i"
8642. .9358
/gene="eps3i"
/codon_start=1
/transl_table=11
/product="Eps3i"
/protein_id="AAL23733.1"
/db_xref="GI:24473742"
/translation="MMYAKTCFLIIAHNNMGOLKYLIEGLDSQTHDIFVHVDKSKDF
DSYFLDSYTKSLKFRREFIVYMGDFSGOVQVEMFLLEQAYIEGIDYHIIISGADLPL

gene
CDS
9374..10474
/gene="eps3J"
9374..10474
/gene="eps3J"
/codon_start=1
/transl_table=11
/product="Eps3J"
/protein_id="AAL23734.1"
/db_xref="GI:24473743"
/translation="MKILQIPGGLSPGDSINSFIVEYVTAMDSENDIRLATNNAE
STQRYVSGCEVSIPIYKONIKYFPLFRYIAEKIDIVHGSASIMSYELAA
KAGCKVRIASHNNTICENORVDQILRPVDDKLYTSCFCGCGAGKMLPGRNFIYIP
NRSLLKRYENQKRLQYRKRLHIPPDALVIGVGRNEGRHKYLIQIYERHKIY
KSYLVLTIGTDVTVKQOYVTELENYVIFIGAVNNVPDLSAFVMDLPSLYEGIP
SVVIEQISGLPCLISDITNECKITSLVFESIETTPETWADIVELKIEDRNIND
KIFKEVDAGYDIDEDAKKLKELYSIYSPFK"
10496..11166
/gene="eps3K"
/pseudo
/note="Eps3K; putative glucosyltransferase"
/pseudo
CDS
gene
Query Match 0.4%; Score 28; DB 1; Length 17468;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1818 TTTTATGATTTTGTACAGTAGAAG 1845
DB 8126 TTTTATGATTTTGTACAGTAGAAG 8153
RESULT 4
LOCUS AX283715 17468 bp DNA linear PAT 20-NOV-2001
DEFINITION Sequence 9 from Patent WO0179500.
ACCESSION AX283715
VERSION AX283715.1 GI:17044426
KEYWORDS
SOURCE Streptococcus thermophilus
ORGANISM Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Rallu, F., Besancon-Yoshpe, I., Fremaux, C., Mengaud, J. and Renault, P.
TITLE Streptococcus thermophilus operons involved in exopolysaccharide (e
JOURNAL ps) synthesis
PATENT: WO 0179500-A 9 25-OCT-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR) ;
COMPAGNIE GRAVAIS-DANONE (FR) ; Rhodia Chimie (FR)
FEATURES
source location/Qualifiers
1..17468
/organism="Streptococcus thermophilus"
/mol_type="unassigned DNA"
/db_xref="taxon:1308"
ORIGIN
Query Match 0.4%; Score 28; DB 6; Length 17468;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1818 TTTTATGATTTTGTACAGTAGAAG 1845
DB 8126 TTTTATGATTTTGTACAGTAGAAG 8153
RESULT 5
LOCUS CO963297 16032 bp DNA linear PAT 06-DEC-2004

DEFINITION Sequence 1 from Patent WO2004085607.
ACCESSION CO963297
VERSION CO963297.1 GI:56403543
KEYWORDS
SOURCE Streptococcus thermophilus
ORGANISM Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Horvath, P., Manoury, E., Huppert, S. and Fremaux, C.
TITLE Texturizing lactic bacteria
JOURNAL Patent: WO 2004085607-A 1 07-OCT-2004;
Rhodia Chimie (FR)
FEATURES
source location/Qualifiers
1..16032
/organism="Streptococcus thermophilus"
/mol_type="unassigned DNA"
/db_xref="taxon:1308"
ORIGIN
Query Match 0.4%; Score 27; DB 6; Length 16032;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1807 TGGGAGATATTTTATGATTTTGT 1833
DB 7076 TGGGAGATATTTTATGATTTTGT 7102
RESULT 6
LOCUS CR931678 19798 bp DNA linear BCT 30-JUN-2005
DEFINITION Streptococcus pneumoniae strain 485/61 (serotype 19f).
ACCESSION CR931678
VERSION CR931678.1 GI:68643444
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 19798)
AUTHORS Bentley, S.D., Aanensen, D., Mavroidi, A., Saunders, D.,
Bentley, S.D., Aanssen, D., Collins, M., Donaghy, K., Harris, D., Kallio, M.S.,
Rabinowitz, E., Collins, M., Donaghy, K., Harris, D., Kallio, M.S.,
Murphy, L., Quail, M.A., Samuel, G., Skovsted, I.C., Barrett, B.G.,
Reeves, P., Parkhill, J. and Spratt, B.G.
TITLE Genetic analysis of the capsular biosynthetic locus from all 90
serotypes of Streptococcus pneumoniae
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19798)
AUTHORS Bentley, S.D.
TITLE Direct Submision
JOURNAL Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
COMMENT NOTE: This sequence was generated from a PCR product representing
the region from dexB to alfa and is not necessarily responsible for
the expressed capsule serotype. For a detailed description of how
CDS products were predicted see the associated publication.
FEATURES
source location/Qualifiers
1..19798
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="485/61"
/db_xref="taxon:1313"
1..134
/gene="dexB"
/locus_tag="SPC19F_0001"
1..134
/gene="dexB"
/locus_tag="SPC19F_0001"
/codon_start=3
/transl_table=11
/protein_id="CAI33688.1"


```

gene
  /db_xref="GI:68643445"
  /translation="SNEBQDLTVEGKVKSVLIENITLAQEFVEKQILVPMWDAFVEILL"
  /gene="aliB"
  /locus_tag="SPC19F_0002"
  /pseudo
  293..461
  /gene="aliB"
  /locus_tag="SPC19F_0002"
  /pseudo
  293..461
  /locus_tag="SPC19F_0002"
  /pseudo
  1107..2384
  /locus_tag="SPC19F_0003"
  /locus_tag="SPC19F_0003"
  /note="member of homology group 0108 3"
  /codon_start=1
  /product="putative group II intron protein"
  /protein_id="CAI33690.1"
  /db_xref="GI:68643446"
  /translation="MSKLLDKILSRNMLEAVNQVSKNSAGIDGMITENDNYIRQ
  NMRLTKELIKORYKQPVLYKVEIPKPDGIRQLGIPVMDRMIOQAIVQVNSPICEP
  HPSDTSYGRPNSSCEKAILKLELVNDGVEWIVDIDLEKFPDVPQDLMSLVNII
  EBGDTSILIRKYLHSGVILNGORVKTIVGPQGSPLISNIMLNELDEKELRGLR
  FVRVADCVITVGSSEAAKRWYSVSRFTEKRLGLKVMNKTKITRPRELKYLGREGFV
  KSSVDSKSRPHDQSVRFKLLKKTIVQRKMSIDITRIEQLNLSIGKININYSLSGMMK
  SIYASIDELRTRLRVITWKQWRKSRRLMGLKIGVPRWIDKVSQMGDHYQLVQK
  SVLKRAISKPVLEKRGIVSCLDYTLERHALKVS"
  1281..1931
  /locus_tag="SPC19F_0003"
  /note="HMPfam hit to PF00078, RNA-directed DNA polymerase
  (reverse transcriptase), score 1.6e-63"
  /complement(2441..3226)
  /gene="tnp"
  /locus_tag="SPC19F_0004"
  /pseudo
  /complement(2441..3226)
  /gene="tnp"
  /locus_tag="SPC19F_0004"
  /note="member of homology group 0004 74"
  /pseudo
  /codon_start=1
  /transl_table=11
  /product="putative IS630-SpnI transposase"
  /complement(2855..3226)
  /gene="tnp"
  /locus_tag="SPC19F_0004"
  /note="HMPfam hit to PF01710, Transposase, Synchocystis
  PCC 6803, score 9.7e-15"
  /pseudo
  3475..4920
  /gene="wzg"
  /locus_tag="SPC19F_0005"
  3475..4920
  /gene="wzg"
  /locus_tag="SPC19F_0005"
  /note="member of homology group 0000 90"
  /codon_start=1
  /transl_table=11
  /product="integral membrane regulatory protein Wzg"
  /protein_id="CAI33692.1"
  /db_xref="GI:68643447"
  /translation="MSRRPFKRSRSQVRKRSVNIIVLTITVLLVCFLLFLIKYNIIAF
  .
```

```

  RYINIVTATVIALVAGLLITVKKAEKFTILVPSILVSVSLFAVQOQFVGLTNR
  INATNSISYISIAVADSDIENVQTSTAPRTGNTENMIQKILAIKSONDILT
  VNQSSSYLAAYKSLIAGETKAIVLNVSVENITLESYPYASKIKITVKGFKVYEA
  KTSKQSFNINIVSGIDTYGPISSVSRSDVNITMTNRPKILITTPRDADVPVADG
  GNNQDKLTNAGIYGVDSIHTELVYGVDTNYYVRLNFTSFLKVIDLGGVADVANDQ
  EFSALHGKHFPGVGNVDSROALQFVRYERSLADGDRGNQGVAVAILOKLTST
  EVLKYSKSIILQGDQSLQTMNPIETMIDLVTQGLESGGNKYKNSODLKGTRMDLPSY
  AMPDSNLVMEIDSSILAVKAIQDVNEGR"
  3475..3591
  /gene="wzg"
  /locus_tag="SPC19F_0005"
  /note="Signal peptide predicted for SPC1032 by SignalP 2.0
  HMM (Signal peptide probability 0.999) with cleavage site
  probability 0.540 between residues 39 and 40"
  join(3523..3582,3610..3678,3697..3765)
  /gene="wzg"
  /locus_tag="SPC19F_0005"
  /note="3 probable transmembrane helices predicted for
  SPC1032 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
  3682..4026
  /gene="wzg"
  /locus_tag="SPC19F_0005"
  /note="HMPfam hit to PF02916, DNA polymerase processivity
  factor, score 1.6e-64"
  4204..4647
  /gene="wzg"
  /locus_tag="SPC19F_0005"
  /note="HMPfam hit to PF03816, Cell envelope-related
  transcriptional attenuator, score 3.2e-70"
  4922..5653
  /gene="wzh"
  /locus_tag="SPC19F_0006"
  4922..5653
  /gene="wzh"
  /locus_tag="SPC19F_0006"
  /BC_number="3.13.48"
  /note="member of homology group 0001 90"
  /codon_start=1
  /transl_table=11
  /product="protein-tyrosine phosphatase Wzh"
  /protein_id="CAI33693.1"
  /db_xref="GI:68643448"
  /translation="MIDHSHTVPVDDGPKRSRESKALIAESYRGVRTVSTSHR
  KMFETPEKIAENFLQVREINKEVADDLVAYGAEIYITTDALKEKKEIPLTANDS
  RYALIEFSMTSTYRQIHGELSNIMLGITPVIAHLERDALENNEKRRELLIDMGCTY
  QINSYHVSKEPFGEKRYKPKRRARYFLERDLVHVASDMNLDLRPPYMOAYDIIA
  KKYGAKVKEELFVDNPRKIIMDQLI"
  5234..5548
  /gene="wzh"
  /locus_tag="SPC19F_0006"
  /note="HMPfam hit to PF02811, PHP, C-terminal, score
  1.5e-10"
  5662..6354
  /gene="wzd"
  /locus_tag="SPC19F_0007"
  5662..6354
  /gene="wzd"
  /locus_tag="SPC19F_0007"
  /note="member of homology group 0002 90"
  /codon_start=1
  /transl_table=11
  /product="capsular polysaccharide biosynthesis protein
  Wzd"
  /protein_id="CAI33694.1"
  /db_xref="GI:68643449"
  /translation="MKEQNTLEIIVDLOLPRALMKKRVITLVAILTSSVAFAFPEVI
  KBFTSMTRIYVNNQSGSKGLTNODDLAGSSLYKDYRTREITLSQVLEEVSDIKLD
  LPPKDLANKIKVTPVDTRIIVSVSVDVPEASRIANLREVAQKISTIRVSDVT
  TLEEARPATSPSSPNIKRSTLIVFLAGVIGTSVIVILIELDTRVRRPKDIDDTLMT
  LIGIVNLNKLK"
  5680..6096
  /gene="wzd"
  /locus_tag="SPC19F_0007"
  .
```

```

/ncseq="HMMFam hit to PF02706, Lipopolysaccharide
biosynthesis, score 3.1e-52"
join(S728..5793,6196..6264)
/genes="wzd"
/locus_tag="SPC19F_0007"
/notes="2 probable transmembrane helices predicted for
SPC1034 by TMHMM2.0 at aa 23-44 and 179-201"
6364..7053
/genes="wze"
/locus_tag="SPC19F_0008"
6364..7053
/genes="wze"

Query Match          0.4%; Score 27; DB 1; Length 19798;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1077 GACGAGTTACCCAGCTTTTATATGTT 1103
Db      8141 GACGAGTTACCCAGCTTTTATATGTT 8167

LOCUS      STH289861      21562 bp      DNA      linear      BCT 15-APR-2005
DEFINITION Streptococcus thermophilus eps locus.
ACCESSION AJ289861
VERSION AJ289861.1 GI:18076389
KEYWORDS   deo gene; deo protein; Eps 2; Eps 3; Eps 4; eps10 gene; Eps10
protein; eps11 gene; Eps11 protein; eps12 gene; Eps12 protein;
eps14 gene; Eps14 protein; eps15 gene; Eps15 protein; eps16 gene;
Eps16 protein; eps2 gene; eps3 gene; eps4 gene; eps5 gene; Eps5
protein; eps9 gene; Eps9 protein; epsA gene; EpsA protein; epsB
gene; EpsB protein; epsC gene; EpsC protein; epsd gene; Epsd
protein; epsE gene; EpsE protein; ORF1; ORF6; pseudogene.
Streptococcus thermophilus
Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE   1
AUTHORS   Pluvinet A., Chartron-Bourgoign F., Morel C. and Decaris B.
TITLE     Implication of horizontal transfers in the chimeric structure of
the eps locus of Streptococcus thermophilus IFe757
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 21562)
AUTHORS   Pluvinet A.
TITLE     Direct Submission
JOURNAL   Submitted (18-MAY-2000) Pluvinet A., Laboratoire de Genetique et
Microbiologie, Universite Henri Poincare - Faculte des Sciences,
B.P. 239, Vandoeuvre-lès-Nancy 54506, FRANCE

FEATURES             source
     source          1..21562
                        /organism="Streptococcus thermophilus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:1308"
                        <1..465
                        /note="ORF1"
                        /codon_start=1
                        /transl_table=11
                        /product="hypothetical protein"
                        /protein_id="CAC82000.1"
                        /db_xref="GI:18076390"
                        /db_xref="UniProt/TREMBL:Q8VM12"
                        /translation="DIYIVGVCLGQLSLVAYGKYSPEKRIIGVPLITLTDGLAD
                        EHIDFGSTLTGTHMGHGMGSLDKAVILASKSKYPIIORSKHYAFQAHLEFDLE
                        AIDLIADGESEHLXQCKQIDPVTPEQLAHNHYSQMKCLPAFLDSLTQI"
                        800..1809
                        /gene="deoB"
                        /gene="deoD"
                        800..1510
                        /gene="deoD"
                        /codon_start=1
                        /evidence=not_experimental
                        /transl_table=11

```

/product="Deod protein"
 /protein_id="CAC82001.1"
 /db_xref="GI:18076391"
 /db_xref="GOA:O8VM11"
 /db_xref="InterPro:IPR000845"
 /db_xref="InterPro:IPR004402"
 /db_xref="UniProt/TREMBL:O8VM11"
 /translaton="MSIHISAKOGDIAFSTIYARELLIDYGVKILRGTAAGSLANENHRE
 MFGYITGKGRISYMGTMGMGMPDSIYARELLIDYGVKILRGTAAGSLANENHRE
 LVLAQAAATNSNIIIRNDMPQYDFPOLANFNILDAVAHIAKNFGMTTHGVANLSSDVEY
 SNYEPKNIELIKMGVKAIVEMEAALYYLAAQHVDAALAMTISDSLVNPDPTTAER
 ONTPDMKVKITLIAD"
 1744. 1764
 /gene="deod"
 1781. 1786
 /gene="deod"
 1804. 1809
 /gene="deod"
 1853. 3307
 /gene="epsa"
 1853. 3307
 /gene="epsa"
 /function="putative exopolysaccharide regulation"
 /codon_start=1
 /evidence-not experimental
 /transl_table=1
 /product="Epsa protein"
 /protein_id="CAC82002.1"
 /db_xref="GI:18076392"
 /db_xref="GOA:O56038"
 /db_xref="InterPro:IPR004190"
 /db_xref="InterPro:IPR004474"
 /db_xref="UniProt/TREMBL:Q56038"
 /translaton="MSSRTNKRQKHTNSGSGWNVNGLITLIVALLFTMNPNF
 LSPRIILIIITGLLVLAISIFLOTKKLP.LVTVVAVLIVSLVIGIFGFMKDI
 TNRNMQATAEVEVMSIVVPKESDIKVSQLTSAQPTVKDKNNEIILMAKKDDKV
 DKVDUDVASYGEADINLKSKSKAMVYLSGSYASILLESVDNSYANLKTITTYIKIKKN
 SNSAQVDSRFENIYISGIDYGPISVSRSDVIIMTVNNTHKILLITTPDAYV
 IPGGAQDYDGLTHAGIYGVTSQTLDEHYGLIDYVARINFTSFLKLDLQGVGV
 HNDQAFTEKEDPVGVDIOMNSEQALGFVERVNLDEGDNDRGKNDKRVSAIINKTA
 SLKTSVNTSIVNNLQDSVOTNMSLINTINLANLTQLESGSKFTVTSQAVGTGSTGL
 ISYMPNSISLTMKLDSVSBSAQAIKLMEEK"
 3308. 4039
 /gene="epsb"
 3308. 4039
 /gene="epsb"
 /codon_start=1
 /evidence-not experimental
 /transl_table=1
 /product="Epsb protein"
 /protein_id="CAC82003.1"
 /db_xref="GI:18076393"
 /db_xref="GOA:O56039"
 /db_xref="InterPro:IPR004013"
 /db_xref="UniProt/TREMBL:Q56039"
 /translaton="MIDWHSIVPDVDDGPEPTLEESLDLIGESYAQGVKRIVSTSHRR
 KGMETPDSKIFANFKYKVAEALYEDLTIYVGBELYYTSDIVEKLEKNLIRPMNPT
 OEALEFPARSKMEIHSIGSNVLRAQVTPVAHIERVYDALENADRVREIINNGCVT
 QVNSHUTIKPLDGDKDKVRKRVRFLEKNLVHNVASDMHNLGPRPPFMKAYEIVK
 KNYGSKRAKNI.FIENPKTILENOYL"
 4048. 4740
 /gene="epscc"
 /gene="epscc"
 4048. 4740
 /gene="epscc"
 /function="chain length determinant and exporter"
 /codon_start=1
 /evidence-not experimental
 /transl_table=1
 /product="Epsc protein"
 /protein_id="CAC82004.1"
 /db_xref="GI:18076394"
 /db_xref="GOA:O8VM10"
 /db_xref="InterPro:IPR003856"

gene
4750. .5490
/gene="epsd"
CDS
4750. .5490
/gene="epsd"
/function="chain length determinant and exporter"
/evidence=not_experimental
/codon_start=1
/transl_table=11
/product="Epsd protein"
/protein_id="CAC82005.1"
/db_xref="GI:18076395"
/db_xref="GOA:Q8VM09"
/db_xref="InterPro:IPR005702"
/db_xref="UniProt/TREMBL:Q8VM09"
/translation="MPLKLVKSKVDPAKKEEYNAIRINIQSGAKVIAISVE
AGEGSGTISVNLAISFASVGLTTLIDAKETNSVLSGTFKSNPEYKGISNFLSNAL
NETTIOOTDISGPDVIAAGVDPNPSTLSLONNFRHLMFVARSCYDYIIDPPVGLVI
DAVITAHQADSLVLTGAKIKRRPVTKAQVLVSSGSQGLGVILNKVDMTVDKGYR
GSGYSGYGRKSKSDQKEGHSRAHRRRKG"
gene
5547. .6914
/gene="epsE"
/pseudo
5547. .6914
/gene="epsE"
/pseudo
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="EpsE protein"
/db_xref="PSEUDO:CAC82006.1"
7116. .8132
/gene="epsI"
/pseudo
7116. .8132
/gene="epsI"
/pseudo
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="EpsI protein"
8148. .8597
/gene="eps2"
8148. .8597
/gene="eps2"
/function="glycosyltransferase"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Eps2 protein"
/protein_id="CAC82008.1"
/db_xref="GI:18076396"
/db_xref="GOA:Q8VM08"
/db_xref="InterPro:IPR002376"
/db_xref="UniProt/TREMBL:Q8VM08"
/translation="MKVCLVSSGGHLAHLNLKPPWSRQDRFVTPPKEDARSILK
EGPYEPLTNRNFKNLVRNTFLAKILIKERPDVLISSGAVAVPFYGLGFAKAK
VYIEVFDRIIDKPVVTGKLAVYPVTKPIVOMEEMKTVYPALINLSIF"
8597. .9082
/gene="eps3"
8597. .9082
/gene="eps3"
/function="glycosyltransferase"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Eps3 protein"

```
/protein_id="CAC82009.1"  
/db_xref="GI:18076397"  
/db_xref="GOA:Q8VM07"  
/db_xref="InterPro:IPR007235"  
/db_xref="UniProt/TREMBL:Q8VM07"
```

[illegible]

RESULT 8
 CP000023_09/c
 WPCOMMENT
 Sequence split into 18 fragments LOCUS CP000023 Accession CP000023

Fragment Name	Begin	End
CP000023_00	1	110000
CP000023_01	100001	210000
CP000023_02	200001	310000
CP000023_03	300001	410000
CP000023_04	400001	510000
CP000023_05	500001	610000
CP000023_06	600001	710000
CP000023_07	700001	810000
CP000023_08	800001	910000
CP000023_09	900001	1010000
CP000023_10	1000001	1110000
CP000023_11	1100001	1210000
CP000023_12	1200001	1310000
CP000023_13	1300001	1410000
CP000023_14	1400001	1510000
CP000023_15	1500001	1610000
CP000023_16	1600001	1710000
CP000023_17	1700001	1796846

Continuation (10 of 18) of CP000023 from base 900001 (CP000023) Streptococcus thermophilus

	Query Match	Score	DB	Length
	Best Local Similarity	100.0%	Prod. No. 1.9'	
Matches	27	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
QY	1807	TCGGGAGATTTTAAATGATTTTGT	1853	
Db	75539	TCGGGAGATTTTAAATGATTTTGT	75513	

RESULT 9	AC121782	16532 bp	DNA	linear	ROD 11-NOV-2003
LOCUS	AC121782				
DEFINITION	Mus musculus BAC clone RP23-367M7 from chromosome 3, complete				
ACCESSION	AC121782				
VERSION	AC121782.3	GI:23499668			
KEYWORDS	HTG.				
SOURCE	Mus musculus				
ORGANISM	Mus musculus				

REFERENCE	1 (bases 1 to 165328)
AUTHORS	Trani, L., Spalding, L. and Doebber, A.
TITLE	The sequence of Mus musculus BAC clone RP23-367M7
JOURNAL	Unpublished (2001)
REFERENCE	2 (bases 1 to 165328)
AUTHORS	Wilson, R.
TITLE	Sequencing of Mus musculus
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 165328)
AUTHORS	McPherson, J.D. and Waterston, R.H.
TITLE	Direct Submission

JOURNAL Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 165328)

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 165328)

TITLE McPherson,J.D. and Waterston,R.H.

JOURNAL Submitted (04-OCT-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 6 (bases 1 to 165328)

AUTHORS Wilson,R.

TITLE Direct Submission

JOURNAL Submitted (11-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Oct 4, 2002 this sequence version replaced gi:22475640.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu>

Contact: submissions@wustl.wustl.edu

----- Summary Statistics -----

Center project name: M_BA0367M07

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minko Tareno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC087062.

Location/Qualifiers

1.165328

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="3"

/map="3"

/clone="RP23-367M7"

/clone_1ib="RPCI-23"

435.492

repeat_region /rpt_family="ERV1"

518.770

repeat_region /rpt_family="MaLR"

771.962

repeat_region /rpt_family="B2"

963.1011

repeat_region

/rpt_family="MaLR"

1057.1115

repeat_region /rpt_family="ERV1"

2284.2416

repeat_region /rpt_family="Alu"

2559.2687

repeat_region /rpt_family="B4"

2576.2805

repeat_region /rpt_family="MaLR"

2715.2788

repeat_region /rpt_family="MaLR"

2808.2855

repeat_region /rpt_family="B4"

2959.3016

repeat_region /rpt_family="Alu"

3050.3254

repeat_region /rpt_family="B2"

3257.3389

repeat_region /rpt_family="Alu"

3547.3694

repeat_region /rpt_family="MaLR"

4715.4861

repeat_region /rpt_family="Alu"

4987.5182

repeat_region /rpt_family="B2"

5189.5325

repeat_region /rpt_family="Alu"

5580.5628

repeat_region /rpt_family="ERV1"

5629.5698

repeat_region /rpt_family="ID"

5703.5799

repeat_region /rpt_family="Alu"

7794.7841

repeat_region /rpt_family="Alu"

7996.8146

repeat_region /rpt_family="B2"

8159.8213

repeat_region /rpt_family="ID"

8691.8728

repeat_region /rpt_family="B4"

8729.8924

repeat_region /rpt_family="B2"

8735.8807

repeat_region /product="tRNA-Ser"

/note="Likely pseudogene (HMM Sc=36.14 / Sec struct Sc=-11.21)"

8925.8999

repeat_region /rpt_family="B4"

9000.9129

repeat_region /rpt_family="Alu"

9130.9206

repeat_region /rpt_family="B4"

9143.9256

repeat_region /rpt_family="B2"

9363.9418

repeat_region /rpt_family="Alu"

9419.9499

repeat_region /rpt_family="B4"

9432.9555

repeat_region /rpt_family="Alu"

9673.9762

repeat_region /rpt_family="B2"

9778.9927

repeat_region /rpt_family="Alu"

11712.11886

repeat_region /rpt_family="L1"

12516.12594

repeat_region /rpt_family="Alu"

12668.12872

repeat_region /rpt_family="B2"

13976.14053

repeat_region /rpt_family="B4"

```

repeat_region      14062..14332
                    /rpt_family="MALR"
repeat_region      14325..14456
                    /rpt_family="MALR"
repeat_region      14470..14668
                    /rpt_family="B2"
repeat_region      14783..14929
                    /rpt_family="Alu"
repeat_region      14973..15095
                    /rpt_family="Alu"
repeat_region      15107..15273
                    /rpt_family="B2"
repeat_region      15307..15495
                    /rpt_family="B2"
repeat_region      15497..15673
                    /rpt_family="B2"
repeat_region      15674..15738
                    /rpt_family="MALR"
repeat_region      15761..15937
                    /rpt_family="B2"
repeat_region      15939..15982
                    /rpt_family="B4"
repeat_region      16672..17056
                    /rpt_family="MALR"
repeat_region      17797..18071
                    /rpt_family="B4"
repeat_region      18144..18216
                    /rpt_family="ID"
repeat_region      18361..18487
                    /rpt_family="Alu"
repeat_region      18523..18702
                    /rpt_family="B2"

```

Query Match 0.4%; Score 27; DB 9; Length 165328;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5723 AGAAGATTACTTTTAACTGCATTA 5749
 Db 93500 AGAAGATTACTTTTAACTGCATTA 93526

```

RESULT 10          183094 bp   DNA   linear   ROD 18-SEP-2003
AC087903/c         Mus musculus strain C57BL/6J clone rp23-4p7 map 3, complete
LOCUS              AC087903
DEFINITION         sequence.
ACCESSION          AC087903
VERSION            AC087903.26 GI:34849938
KEYWORDS           HTG.
SOURCE             Mus musculus (house mouse)
ORGANISM           Mus musculus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                  Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            1 (bases 1 to 183094)
TITLE             Mus musculus BAC Clone rp23-4p7
JOURNAL            Unpublished
                  2 (bases 1 to 183094)
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            2 (bases 1 to 183094)
TITLE             Direct Submission
JOURNAL            Submitted (06-FEB-2001) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            3 (bases 1 to 183094)
TITLE             Direct Submission
JOURNAL            Submitted (19-OCT-2002) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            4 (bases 1 to 183094)
TITLE             Direct Submission
JOURNAL            Submitted (19-OCT-2002) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA

```

```

TITLE             Direct Submission
JOURNAL            Submitted (22-OCT-2002) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            5 (bases 1 to 183094)
TITLE             Direct Submission
JOURNAL            Submitted (15-NOV-2002) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            6 (bases 1 to 183094)
TITLE             Direct Submission
JOURNAL            Submitted (27-NOV-2002) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            7 (bases 1 to 183094)
TITLE             Direct Submission
JOURNAL            Submitted (10-DEC-2002) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            8 (bases 1 to 183094)
TITLE             Direct Submission
JOURNAL            Submitted (18-SEP-2003) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA
COMMENT            On Sep 18, 2003 this sequence version replaced gi:26291169.
                  ----- Genome Center
                  Center: Department Of Chemistry And Biochemistry
                  The University Of Oklahoma
                  Center code:UOKNOR

```

FEATURES
 source Location/Qualifiers

```

1..183094
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /map="3"
  /clone="rp23-4p7"
  /clone_11b="RPCL - 23 Female (C57BL/6J) Mouse BAC library"

```

ORIGIN

Query Match 0.4%; Score 27; DB 9; Length 183094;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5723 AGAAGATTACTTTTAACTGCATTA 5749
 Db 172159 AGAAGATTACTTTTAACTGCATTA 172133

```

RESULT 11          224451 bp   DNA   linear   ROD 18-SEP-2003
AC087062/c         Mus musculus strain C57BL/6J clone rp23-15p14 map 3, complete
LOCUS              AC087062
DEFINITION         sequence.
ACCESSION          AC087062
VERSION            AC087062.25 GI:34849932
KEYWORDS           HTG.
SOURCE             Mus musculus (house mouse)
ORGANISM           Mus musculus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                  Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            1 (bases 1 to 224451)
TITLE             Mus musculus BAC Clone rp23-15p14
JOURNAL            Unpublished
                  2 (bases 1 to 224451)
REFERENCE          Do,T., Do,A. and Roe,B.A.
AUTHORS            2 (bases 1 to 224451)
TITLE             Direct Submission
JOURNAL            Submitted (19-OCT-2002) Department Of Chemistry And Biochemistry,
                  The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                  OK 73019, USA

```

TITLE Direct Submission
JOURNAL Submitted (02-DEC-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (17-APR-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 5 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 6 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 7 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 8 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 9 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 10 (bases 1 to 224451)
AUTHORS Do,T., Do,A. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Sep 18, 2003 this sequence version replaced gi:25046365.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
source 1. 224451
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="3"
/clone="rp23-15p14"
/clone_lib="RPC1 - 23 Female (C57BL/6J) Mouse BAC Library"
ORIGIN
Query Match 0.4%; Score 27; DB 9; Length 224451;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5723 AGAAGATTACTTTTATCTGCATTA 5749
Db 12017 AGAAGATTACTTTTATCTGCATTA 11991
RESULT 12
LOCUS CO645816 447 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 2773 from Patent WO0234771.
ACCESSION CO645816
VERSION CO645816.1 GI:41682634
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 2773 02-MAY-2002;
Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source 1. 447
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
ORIGIN
Query Match 0.4%; Score 26; DB 6; Length 447;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1755 TTTATTGTTCAAGTGGGAAGAAATGAA 1780
Db 382 TTTATTGTTCAAGTGGGAAGAAATGAA 407
RESULT 13
LOCUS AX607685 450 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 5614 from Patent WO02092818.
ACCESSION AX607685
VERSION AX607685.1 GI:28403217
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Glaeser,P., Ruhnok,C., Chevalier,F., Frangoul,L., Lalioui,L.,
Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Tien-Cuot,P. and
Kunst,F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 5614 21-NOV-2002;
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source 1. 450
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
ORIGIN
Query Match 0.4%; Score 26; DB 6; Length 450;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1755 TTTATTGTTCAAGTGGGAAGAAATGAA 1780
Db 382 TTTATTGTTCAAGTGGGAAGAAATGAA 407

RESULT 14
AY257676 688 bp DNA linear BCT 01-JAN-2005
LOCUS Streptococcus agalactiae strain AL-96-1683 capsular polysaccharide
DEFINITION biosynthesis gene cluster, partial sequence.
ACCESSION AY257676
VERSION AY257676.1 GI:32364285
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 688)
AUTHORS Kong, F., Gilbert, G.L., Gottschalk, M. and Martinez, G.
TITLE Genetic population structure of Canadian bovine Streptococcus
agalactiae (group B streptococcus, GBS) isolates: Further study by
a GBS genotyping system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 688)
AUTHORS Kong, F., Gilbert, G.L., Gottschalk, M. and Martinez, G.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-2003) CIDMLS, ICPMR, Westmead Hospital, Darcy
Road, Sydney, NSW 2145, Australia
Location/Qualifiers
FEATURES
source 1..688
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/strain="AL-96-1683"
/db_xref="taxon:1311"
<1..60
/gene="cpsE"
<1..60
/gene="cpsF"
/codon_start=1
/transl_table=11
/product="CpsE"
/protein_id="AAP80242.1"
/db_xref="GI:32364286"
/translation="DIKILITLKVLLGTGAK"
84..533
/gene="cpsF"
84..533
/gene="cpsF"
/codon_start=1
/transl_table=11
/product="CpsF"
/protein_id="AAP80243.1"
/db_xref="GI:32364287"
/translation="MKICLVSSGGHLAHLNLKPIWEKEDRFVTFDEKARSILRE
EIVYHCFPTNRNVKNLVKNTLLAFKLRKRPDVIISSGAAVAAPFFYIGKLGCKT
VIVLVPDRIDKPLTGLKLVYPVDKPIVQMBEMKVPKAINLGIF"
533..>688
/gene="cpsG"
533..>688
/gene="cpsG"
/codon_start=1
/transl_table=11
/product="CpsG"
/protein_id="AAP80244.1"
/db_xref="GI:32364288"
/translation="MIFVTVGTHGQGFNRLIKEVDRLKGTGAIDQEVFIQTGYSDFEP
ONCOWSKP"
ORIGIN
Query Match 0.4%; Score 26; DB 1; Length 688;
Best Local Similarity 100.0%; Pred.No.17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1755 TTTATGTTCAAGTGGGAAGAATGAA 1780
|||||
Db 465 TTTATGTTCAAGTGGGAAGAATGAA 490

RESULT 15
AY257677 688 bp DNA linear BCT 01-JAN-2005
LOCUS Streptococcus agalactiae strain RF-96-1993 capsular polysaccharide
DEFINITION biosynthesis gene cluster, partial sequence.
ACCESSION AY257677
VERSION AY257677.1 GI:32364289
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 688)
AUTHORS Kong, F., Gilbert, G.L., Gottschalk, M. and Martinez, G.
TITLE Genetic population structure of Canadian bovine Streptococcus
agalactiae (group B streptococcus, GBS) isolates: Further study by
a GBS genotyping system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 688)
AUTHORS Kong, F., Gilbert, G.L., Gottschalk, M. and Martinez, G.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-2003) CIDMLS, ICPMR, Westmead Hospital, Darcy
Road, Sydney, NSW 2145, Australia
Location/Qualifiers
FEATURES
source 1..688
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/strain="RF-96-1993"
/db_xref="taxon:1311"
<1..60
/gene="cpsE"
<1..60
/gene="cpsE"
/codon_start=1
/transl_table=11
/product="CpsE"
/protein_id="AAP80245.1"
/db_xref="GI:32364290"
/translation="DIKILITLKVLLGTGAK"
84..533
/gene="cpsF"
84..533
/gene="cpsF"
/codon_start=1
/transl_table=11
/product="CpsF"
/protein_id="AAP80246.1"
/db_xref="GI:32364291"
/translation="MKICLVSSGGHLAHLNLKPIWEKEDRFVTFDEKARSILRE
EIVYHCFPTNRNVKNLVKNTLLAFKLRKRPDVIISSGAAVAAPFFYIGKLGCKT
VIVLVPDRIDKPLTGLKLVYPVDKPIVQMBEMKVPKAINLGIF"
533..>688
/gene="cpsG"
533..>688
/gene="cpsG"
/codon_start=1
/transl_table=11
/product="CpsG"
/protein_id="AAP80247.1"
/db_xref="GI:32364292"
/translation="MIFVTVGTHGQGFNRLIKEVDRLKGTGAIDQEVFIQTGYSDFEP
ONCOWSKP"
ORIGIN
Query Match 0.4%; Score 26; DB 1; Length 688;
Best Local Similarity 100.0%; Pred.No.17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1755 TTTATGTTCAAGTGGGAAGAATGAA 1780
|||||
Db 465 TTTATGTTCAAGTGGGAAGAATGAA 490

Tue Dec 27 11:31:34 2005

us-09-767-041-9.olg.rge

Page 17

Search completed: December 26, 2005, 00:24:54
Job time : 22528 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 11:35:51 ; Search time 2505 Seconds
(without alignments)
18602.607 Million cell updates/sec

Title: US-09-767-041-9

Sequence: 1 atcgcccaacgaattgcga.....gcaaacagcaaaagaagctt 6992

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 4996997 seqs, 3332346308 residues

Word size: 0

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

N_Geneseq_21: *
1: geneseqn1980s: *
2: geneseqn1980s: *
3: geneseqn2000s: *
4: geneseqn2001bs: *
5: geneseqn2001bs: *
6: geneseqn2002as: *
7: geneseqn2002bs: *
8: geneseqn2003as: *
9: geneseqn2003bs: *
10: geneseqn2003cs: *
11: geneseqn2003ds: *
12: geneseqn2004as: *
13: geneseqn2004bs: *
14: geneseqn2005s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6992	100.0	6992	3	Aaz60930 Nucleotide
2	611	8.7	26281	3	Aaz60929 Nucleotide
3	28	0.4	17468	6	ABA01441 Streptoc
4	27	0.4	16032	13	ADS13200 S. thermo
5	26	0.4	447	6	ABN67430 Streptoc
6	26	0.4	450	13	ADV84473 Streptoc
7	26	0.4	2226	11	ADM79774 Group B S
8	26	0.4	6865	6	ABK90550 Beta1.3-g
9	26	0.4	17276	8	ACA64723 Streptoc
10	26	0.4	17276	10	ADF43363 Streptoc
11	26	0.4	17276	14	AEA03034 Streptoc
12	26	0.4	25020	12	ADO40235 S. agalac
13	26	0.4	95596	13	ADV87741 Streptoc
14	26	0.4	95596	13	ADV87741 Streptoc
15	26	0.4	110000	6	ABN71527_11 Streptoc
16	26	0.4	110000	13	ADO81204_12 Streptoc
17	26	0.3	127145	13	ADQ80254 Human pro
18	23	0.3	734	5	ABV15535 Human pro
19	23	0.3	5139	3	AAV70139 Plasmodiu

20	23	0.3	5818	4	AAS46636 Tumour su
21	23	0.3	6686	6	ABL32460 Human imm
22	23	0.3	7004	4	AAS45413 Chemical
23	23	0.3	7004	6	ABK28266 DNA trans
24	23	0.3	8896	6	ABN80326 Human che
25	23	0.3	10688	4	ABL03406 Human che
26	23	0.3	12135	4	ABL03408 Drosophi
27	23	0.3	20486	6	ABL34610 Human met
28	23	0.3	20486	7	ADS99871 Bilephit
29	23	0.3	80073	2	AATS8840_5 Contnuation (6 of
30	23	0.3	96289	13	ABD33205 Murine ca
31	23	0.3	178024	12	ADQ97721 Human can
32	23	0.3	471	6	ABN65492 Human can
33	22	0.3	579	6	ADA31881 DNA encod
34	22	0.3	5368	6	ABN80098 Human che
35	22	0.3	6294	6	ABL33054 Human imm
36	22	0.3	7134	6	ABL32482 Human imm
37	22	0.3	8951	6	ABL32795 Human imm
38	22	0.3	9299	6	ABL33725 Human imm
39	22	0.3	9515	4	AAS45452 Chemical
40	22	0.3	9515	6	ABK28307 DNA trans
41	22	0.3	9515	6	ABN80244 Human che
42	21	0.3	475	8	ABX53771 Bovine ES
43	21	0.3	488	9	ACH36504 Human end
44	21	0.3	575	5	ABA11020 Human ner
45	21	0.3	692	5	ABA14376 Human ner

ALIGNMENTS

RESULT 1	
AAZ60930	
ID	AAZ60930 standard; DNA; 6992 BP.
XX	
XX	AAZ60930;
AC	
XX	
DT	30-MAY-2000 (first entry)
DE	Nucleotide sequence of a capsular gene cluster of S. suis serotype 1.
XX	
KW	Capsular gene cluster; serotype 1; polysaccharide biosynthesis;
KW	capsular component; antigen; regulation; chain length determination;
KW	complement-mediated opsonophagocytosis; serotype-specific detection;
KW	antigen; vaccine; Streptococcal disease; CpsII, CpsI, CpsII; CpsII;
XX	
OS	Streptococcus suis.
XX	
EH	Location/Qualifiers
FT	2..1367
FT	/*tag= a
FT	/note= "CpsII; encodes AAY68970"
FT	1374..1823
FT	/*tag= b
FT	/note= "CpsII; encodes AAY68971"
FT	1823..2317
FT	/*tag= c
FT	/note= "CpsII; encodes AAY68972"
FT	3036..4202
FT	/*tag= d
FT	/note= "CpsII; encodes AAY68973; the nucleotides encoding
FT	amino acid 213 are not given"
FT	4195..5163
FT	/*tag= e
FT	/note= "CpsII; encodes AAY68974"
FT	5172..6143
FT	/*tag= f
FT	/transl_except= (pos: 5634..5636, aa: Xaa)
FT	/transl_except= (pos: 5637..5638, aa: Xaa)
FT	/note= "CpsII; xaa is an unspecified amino acid; encodes
FT	AAY68975; nucleotides 5328-5330 encode an amino acid that
FT	is not included in the protein sequence"

FT CDS 6156..6990
FT /tag= 9
FT /note= "CpsA; encodes AAY68976; no termination codon
FT given"
PN WO200005378-A2.
XX 03-FEB-2000.
XX 19-JUL-1999; 99WO-NL000460.
XX 22-JUL-1998; 98BP-00202465.
XX 22-JUL-1998; 98BP-00202467.
XX (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX Smith HE;
XX WPI; 2000-195104/17.
XX P-PSDB; AAY68970, AAY68971, AAY68972, AAY68973, AAY68974, AAY68975,
XX AAY68976.
XX New nucleic acid containing the capsular gene cluster of Streptococcus
XX suis, used for serotype-specific detection and to generate antigens or
XX mutants for vaccination.
XX Claim 5; Fig 4; 144pp; English.
XX The present sequence represents the capsular gene cluster of
XX Streptococcus suis serotype 1. The genes in this cluster are involved in
XX polysaccharide biosynthesis of capsular components and antigens. The
XX proteins have glycosyltransferase activities (CpsII, CpsIII, CpsIV, CpsV,
XX CpsVI) and CP polymerase activities (CpsH). The capsule confers
XX bacterium resistance to complement-mediated opsonophagocytosis. The gene
XX cluster is used as a source of probes and primers for serotype-specific
XX detection of S. suis and is also useful for recombinant production of the
XX proteins. The proteins are then useful for producing antigens that can be
XX used in vaccines, for controlling or eradicating a Streptococcal disease,
XX in humans or animals, e.g. against S. suis in pigs
XX
XX Sequence 6992 BP; 2513 A; 788 C; 1214 G; 2477 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 6992; DB 3; Length 6992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCCCAACGAATTGGCATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA 60
DB 1 ATGCCCAACGAATTGGCATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA 60
QY 61 CAAGTCATATACCAAAATGCTGATTTAAATGCTGCAATTTTATCATATGATGTTTC 120
DB 61 CAAGTCATATACCAAAATGCTGATTTAAATGCTGCAATTTTATCATATGATGTTTC 120
QY 121 ATTATTTGCAATTTTATATCTCGTATGCCAGTTGAATTTAGATATAGAAGTATCTGA 180
DB 121 ATTATTTGCAATTTTATATCTCGTATGCCAGTTGAATTTAGATATAGAAGTATCTGA 180
QY 181 TAGAGTTTGAAGAAACATTTAATACTATATATTTTGAATTTTCTTAACGGCAGTAT 240
DB 181 TAGAGTTTGAAGAAACATTTAATACTATATATTTTGAATTTTCTTAACGGCAGTAT 240
QY 241 CATTTTGTGGAGAAATATTTGCACTTCACAGCGTGGCCGTATTTTCAATTTAA 300
DB 241 CATTTTGTGGAGAAATATTTGCACTTCACAGCGTGGCCGTATTTTCAATTTAA 300
QY 301 TAAACCTGTTTGGTATACCTATTTAAGTAATTTAAGCAATTTAAGGATAGCTTTC 360
DB 301 TAAACCTGTTTGGTATACCTATTTAAGTAATTTAAGCAATTTAAGGATAGCTTTC 360
QY 361 TATTTTGCACATCTATCAAAAAAGAGATTTCTAATTTCAACGGCTGAACGATGGGAAA 420
DB 361 TATTTTGCACATCTATCAAAAAAGAGATTTCTAATTTCAACGGCTGAACGATGGGAAA 420

QY 421 ATATGCAAGTTTATTTGATACATTAACAAATTTCAAAAAAATCTGTGATTTGTAG 480
DB 421 ATATGCAAGTTTATTTGATACATTAACAAATTTCAAAAAAATCTGTGATTTGTAG 480
QY 481 TTTTAGGTACGAATATAGTAAATTAATTTATATCATATACCGCTATTTATTTCTGGAAG 540
DB 481 TTTTAGGTACGAATATAGTAAATTAATTTATATCATATACCGCTATTTATTTCTGGAAG 540
QY 541 AAGCTATAGAGTTTTCACAAAGGAAGTGTGACACGCTTTTAAATCTTACCAAGTG 600
DB 541 AAGCTATAGAGTTTTCACAAAGGAAGTGTGACACGCTTTTAAATCTTACCAAGTG 600
QY 601 AGTTTGTAGCGTAAAGCAATTTCTGATGATTTTGAATTTGTAGTATGATTAACG 660
DB 601 AGTTTGTAGCGTAAAGCAATTTCTGATGATTTTGAATTTGTAGTATGATTAACG 660
QY 661 TTGATATTAATTCATTCGGTTTATCTGGGTTGAAGAAACAAAAAATCCACGCTAGGTT 720
DB 661 TTGATATTAATTCATTCGGTTTATCTGGGTTGAAGAAACAAAAAATCCACGCTAGGTT 720
QY 721 ACCATAGCATTTGTAACTTTTTCACAAATTTTATAGCTATGATATCATATGATGAAC 780
DB 721 ACCATAGCATTTGTAACTTTTTCACAAATTTTATAGCTATGATATCATATGATGAAC 780
QY 781 GACTTTGGATATCTCGAGCGGTATGTCGGGTATTAATTTGTGATATGATTTCTATT 840
DB 781 GACTTTGGATATCTCGAGCGGTATGTCGGGTATTAATTTGTGATATGATTTCTATT 840
QY 841 TGTATGTTCCATTTATTTCTGATAGATGGTGAAGCGGCTATTTTGTGCAAGAACAGTTG 900
DB 841 TGTATGTTCCATTTATTTCTGATAGATGGTGAAGCGGCTATTTTGTGCAAGAACAGTTG 900
QY 901 GACAGATGAGCAGATATTTTACATTTCTACAAATTTGATGATGATGATGATGATGATG 960
DB 901 GACAGATGAGCAGATATTTTACATTTCTACAAATTTGATGATGATGATGATGATGATG 960
QY 961 AAGGCAAAAAAGACTTGTCTCAGCCAAACAGATGCAAGGTTGGTATTTTAAATAG 1020
DB 961 AAGGCAAAAAAGACTTGTCTCAGCCAAACAGATGCAAGGTTGGTATTTTAAATAG 1020
QY 1021 GAAAAACGATCTAGAAATTAATCTGCAATTTGATGATTTTCTGCAAGAACAGTTAGAG 1080
DB 1021 GAAAAACGATCTAGAAATTAATCTGCAATTTGATGATTTTCTGCAAGAACAGTTAGAG 1080
QY 1081 AGTTACCAAGTTTATATGTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1140
DB 1081 AGTTACCAAGTTTATATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1140
QY 1141 CTACAGTTGATGATTTTGAAGAAATATATCTCTGCTCAAAAGAGACATTTGATTTAAAC 1200
DB 1141 CTACAGTTGATGATTTTGAAGAAATATATCTCTGCTCAAAAGAGACATTTGATTTAAAC 1200
QY 1201 CAGGATTAACAGTCTCTGCAAGTTAGTGTGATGATGATGATGATGATGATGATGATG 1260
DB 1201 CAGGATTAACAGTCTCTGCAAGTTAGTGTGATGATGATGATGATGATGATGATGATG 1260
QY 1261 TAGTTGGTTGACCTTACATATGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1320
DB 1261 TAGTTGGTTGACCTTACATATGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1320
QY 1321 TATTAAGAAGAGTGAATTTTGTGAGAGAGGAAATGAATTAAGTATGAAAG 1380
DB 1321 TATTAAGAAGAGTGAATTTTGTGAGAGAGGAAATGAATTAAGTATGAAAG 1380
QY 1381 TTTGTTGGTGGTCTTCAAGGAGGACATTTGATCTATTTTGTATTTTAAACCGTTT 1440
DB 1381 TTTGTTGGTGGTCTTCAAGGAGGACATTTGATCTATTTTGTATTTTAAACCGTTT 1440
QY 1441 GGAAGGAAGAAAGCTTTTGGGTAACATTTGATTAAGAGATGCAAGAAAGCTTTTGA 1500
DB 1441 GGAAGGAAGAAAGCTTTTGGGTAACATTTGATTAAGAGATGCAAGAAAGCTTTTGA 1500

QY 1501 AGAATGAAAAATGATCATGTTACTTTCACAAATCGCAATCTCATTAATTTAGTGA 1560
DB 1501 AGAATGAAAAATGATCATGTTACTTTCACAAATCGCAATCTCATTAATTTAGTGA 1560
QY 1561 AAAATACTTCTTAAGCTTTCAAAAATTTTACGTGATGGAACCAAGATGTTATTTTCAT 1620
DB 1561 AAAATACTTCTTAAGCTTTCAAAAATTTTACGTGATGGAACCAAGATGTTATTTTCAT 1620
QY 1621 CTGATGCGCGCTGTGCTGCCCTCTTTTACATCGGAAAATCTATTGAGCAAAAGCA 1680
DB 1621 CTGATGCGCGCTGTGCTGCCCTCTTTTACATCGGAAAATCTATTGAGCAAAAGCA 1680
QY 1681 TTTATATGAAATGATTCGAGTTAATAATCTACATTAACCTGAAAACCTAGTTATTC 1740
DB 1681 TTTATATGAAATGATTCGAGTTAATAATCTACATTAACCTGAAAACCTAGTTATTC 1740
QY 1741 CCGTAAACAGATTTTTTATGTTCTCATGAGGAAAGAAAGATATATCTTAATCTTA 1800
DB 1741 CCGTAAACAGATTTTTTATGTTCTCATGAGGAAAGAAAGATATATCTTAATCTTA 1800
QY 1801 TTTAATCTGGGAGATTTTTTATGATTTTTTGTATGATGAGAACTCATGAACAGAGT 1860
DB 1801 TTTAATCTGGGAGATTTTTTATGATTTTTTGTATGATGAGAACTCATGAACAGAGT 1860
QY 1861 TTAATCGATGATTAAGAGATGATTTATTTGAAATGAAATGAAATGAAATGAAATTTCT 1920
DB 1861 TTAATCGATGATTAAGAGATGATTTATTTGAAATGAAATGAAATGAAATGAAATTTCT 1920
QY 1921 ATTTATTCAAACAGATATTTCTGATATATTCAGAAATATTCGAAATGAAATTTCT 1980
DB 1921 ATTTATTCAAACAGATATTTCTGATATATTCAGAAATATTCGAAATGAAATTTCT 1980
QY 1981 CAGTTACAAAGAAATGGAACATATATTAACAAATCGAAGTACTTTTGGCCACGAGG 2040
DB 1981 CAGTTACAAAGAAATGGAACATATATTAACAAATCGAAGTACTTTTGGCCACGAGG 2040
QY 2041 CCGCGCTACTTTATGATTCATTTATTCGAAAGGAAAAACAATTTATTTCTAGACA 2100
DB 2041 CCGCGCTACTTTATGATTCATTTATTCGAAAGGAAAAACAATTTATTTCTAGACA 2100
QY 2101 AAAAAATGATGGAACATGTAATGATCATCAAGTGAAGTTGTGAAGAAATTTTACA 2160
DB 2101 AAAAAATGATGGAACATGTAATGATCATCAAGTGAAGTTGTGAAGAAATTTTACA 2160
QY 2161 AGATTAATATATTTATTTATTAAGAAAATATAGATGATTTGTTGAAAAATTTAGAGT 2220
DB 2161 AGATTAATATATTTATTTATTAAGAAAATATAGATGATTTGTTGAAAAATTTAGAGT 2220
QY 2221 TTTCTAAGCAACTAATCTTACATCAATTAATTTTTTTTGTGAAGATTTAAACAAT 2280
DB 2221 TTTCTAAGCAACTAATCTTACATCAATTAATTTTTTTTGTGAAGATTTAAACAAT 2280
QY 2281 AGTTGAAAAATTTAATGAGATCAAGAAATGAATTAATAAAAAAGATCATTTTGTGA 2340
DB 2281 AGTTGAAAAATTTAATGAGATCAAGAAATGAATTAATAAAAAAGATCATTTTGTGA 2340
QY 2341 TGGCTTATCAATATTTTCTCAAGTTTCTGAGAGGATACAGATATTTATCATCTTCT 2400
DB 2341 TGGCTTATCAATATTTTCTCAAGTTTCTGAGAGGATACAGATATTTATCATCTTCT 2400
QY 2401 CTGAGAGATGACACCATTAAGTTCTCTGAGAAATCTGTATTAATTAATTAATATTC 2460
DB 2401 CTGAGAGATGACACCATTAAGTTCTCTGAGAAATCTGTATTAATTAATTAATATTC 2460
QY 2461 TCAGATTTATATGATTTTCAAAAAGATGAGCAAAAATATTAAGAAAATGAGATATA 2520
DB 2461 TCAGATTTATATGATTTTCAAAAAGATGAGCAAAAATATTAAGAAAATGAGATATA 2520
QY 2521 TGAACGAGTTAAATGTTACAGATTTATTTCTAATATATCAAGAAAAATCTATGATATGT 2580
DB 2521 TGAACGAGTTAAATGTTACAGATTTATTTCTAATATATCAAGAAAAATCTATGATATGT 2580
QY 2581 ACTGTTAGAAATTTTATTAAGAAATGATCGAGCTTTGAAATACATTTTACAAAGATTTGT 2640

DB 2581 ACTGTTAGAAATTTTATTAAGAAATGATCGAGCTTTTGAATCTATTTTACAAAGATTTGT 2640
QY 2641 GTTTATTTAGAAATTAATAAATCATGGCTTAAGATATAGATTTGGTCTTAATGGCTTTCG 2700
DB 2641 GTTTATTTAGAAATTAATAAATCATGGCTTAAGATATAGATTTGGTCTTAATGGCTTTCG 2700
QY 2701 CTTCACATGATTTTGTGCAATCTCTTTATCAATGAAAACGAAACAGCTTATTTATTT 2760
DB 2701 CTTCACATGATTTTGTGCAATCTCTTTATCAATGAAAACGAAACAGCTTATTTATTT 2760
QY 2761 AAGTAATCTAATATGTCAGATGAACCTATTTATACAGCAATTAAGAAAAATATGAAATTT 2820
DB 2761 AAGTAATCTAATATGTCAGATGAACCTATTTATACAGCAATTAAGAAAAATATGAAATTT 2820
QY 2821 TCAAAATGATTTATCTAAATATGGAATTTAAGATTTAAGTGAAGAAAAATCAACATCT 2880
DB 2821 TCAAAATGATTTATCTAAATATGGAATTTAAGATTTAAGTGAAGAAAAATCAACATCT 2880
QY 2881 TCTCTATTTGCTTTACAGATGATCTAATTTGATGAATTTGCTAAATGCAAGAAATTTAGT 2940
DB 2881 TCTCTATTTGCTTTACAGATGATCTAATTTGATGAATTTGCTAAATGCAAGAAATTTAGT 2940
QY 2941 TTTTATTTGCTAAGAAATGTAATAAATGAAATTAATCTAATTTTAAAGAAATTTACT 3000
DB 2941 TTTTATTTGCTAAGAAATGTAATAAATGAAATTAATCTAATTTTAAAGAAATTTACT 3000
QY 3001 AAAAAATTAATATGTTGATTTTGTGAGATATATATGTTTAAATTTAATATATAC 3060
DB 3001 AAAAAATTAATATGTTGATTTTGTGAGATATATATGTTTAAATTTAATATATAC 3060
QY 3061 CGGAATATTTTATTTTAAATGATCTCGTGTGATTTTATTTTCCAGACAAAGATATG 3120
DB 3061 CGGAATATTTTATTTTAAATGATCTCGTGTGATTTTATTTTCCAGACAAAGATATG 3120
QY 3121 TATTTTATTAATTTTATGAAATTTAATTTATTTTCAATATTAATTTTGAATTTTAA 3180
DB 3121 TATTTTATTAATTTTATGAAATTTAATTTATTTTCAATATTAATTTTGAATTTTAA 3180
QY 3181 TAAATTAATAAATAATGAAATTTTATGTTTATTAAGTCTAATATATGTTTGTTCAG 3240
DB 3181 TAAATTAATAAATAATGAAATTTTATGTTTATTAAGTCTAATATATGTTTGTTCAG 3240
QY 3241 TACTGCAAGATGATTTTGTGAATTAATTTTGAAGAAATTAATTTGAGATTTTACTGCTC 3300
DB 3241 TACTGCAAGATGATTTTGTGAATTAATTTTGAAGAAATTAATTTGAGATTTTACTGCTC 3300
QY 3301 CCATTAATTTGATTTATGCAATATGATTAATTAATTTGATTCATTTATTAATTTGAT 3360
DB 3301 CCATTAATTTGATTTATGCAATATGATTAATTAATTTGATTCATTTATTAATTTGAT 3360
QY 3361 ATAAAAATTAATAAATATGATCTTTTATGTTTTTATGTTTTATTTAGTATATCTGCAT 3420
DB 3361 ATAAAAATTAATAAATATGATCTTTTATGTTTTTATGTTTTATTTAGTATATCTGCAT 3420
QY 3421 TGTATATTAATCAAAATGGAAGATATGTTTATTTTGAACACACCTTATGAGATAG 3480
DB 3421 TGTATATTAATCAAAATGGAAGATATGTTTATTTTGAACACACCTTATGAGATAG 3480
QY 3481 ACTATCTTATTAACAGCGGTCAAAAACAGGTTGGCTTATGAACTATCTTACGTTAA 3540
DB 3481 ACTATCTTATTAACAGCGGTCAAAAACAGGTTGGCTTATGAACTATCTTACGTTAA 3540
QY 3541 ATACCACTAATTAATATGTTTCAATTCGTTAATCTTTGACATTTAATAAATTAATATGC 3600
DB 3541 ATACCACTAATTAATATGTTTCAATTCGTTAATCTTTGACATTTAATAAATTAATATGC 3600
QY 3601 AACAATTTTCTGCTGCTGCTGCTTTATACAGATCTATTAAGTGAATGAGATATG 3660
DB 3601 AACAATTTTCTGCTGCTGCTGCTTTATACAGATCTATTAAGTGAATGAGATATG 3660
QY 3661 GTAGTTATGCTAGCAATTTAATTAATGCTTTGATGAGATATATAGTGAAT 3720

Db 3661 GTAGTTTATCGTAGCAATATTAATTATATGCTGTGTATGAGATATATAGTGGAAAAAT 3720
Qy 3721 TTGCTTGAATAAAAAAGCTAATAGTATAATTTGTAAATACCTATTAATTTTAAATACAG 3780
Db 3721 TTGCTTGAATAAAAAAGCTAATAGTATAATTTGTAAATACCTATTAATTTTAAATACAG 3780
Qy 3781 AATTCCTTTACCAAGAAATTTTGGCTGTATTAATTTCTAGAGAAATCAAGTAACGAAGCTA 3840
Db 3781 AATTCCTTTACCAAGAAATTTTGGCTGTATTAATTTCTAGAGAAATCAAGTAACGAAGCTA 3840
Qy 3841 GATTTATTTATTTATCAAGAGAAATTTGATTAAGATATTAAGAAAAAATTTTATTTTGGAT 3900
Db 3841 GATTTATTTATTTATCAAGAGAAATTTGATTAAGATATTAAGAAAAAATTTTATTTTGGAT 3900
Qy 3901 ATGGAATATCCGAATTTGAGTTACGGGAACTTGGCTCGGAATCATTCAGGCTATATAT 3960
Db 3901 ATGGAATATCCGAATTTGAGTTACGGGAACTTGGCTCGGAATCATTCAGGCTATATAT 3960
Qy 3961 CATTTTATTAATATGAGAAATAGTTGGTTGAATTTTACTGATGTTTTCTTTTATATG 4020
Db 3961 CATTTTATTAATATGAGAAATAGTTGGTTGAATTTTACTGATGTTTTCTTTTATATG 4020
Qy 4021 TTATATAAAAAAGTTATGAGTTAATGGGAAAACAGACATATTTTATTTTACATCATAG 4080
Db 4021 TTATATAAAAAAGTTATGAGTTAATGGGAAAACAGACATATTTTATTTTACATCATAG 4080
Qy 4081 CCATATTTTTCATATATGAAACAATAGATCCGATTTATTAATTAATAGTACTATCTTT 4140
Db 4081 CCATATTTTTCATATATGAAACAATAGATCCGATTTATTAATTAATAGTACTATCTTT 4140
Qy 4141 CTTCATATAGTATTTGGAAATATATTAATTTTAAAAAGATATGAGACAAAAATGAAAT 4200
Db 4141 CTTCATATAGTATTTGGAAATATATTAATTTTAAAAAGATATGAGACAAAAATGAAAT 4200
Qy 4201 GATTTAATTTCACTATTTGACCAATTTATATATGTCGAAGATTTCTTGATTAATATGAT 4260
Db 4201 GATTTAATTTCACTATTTGACCAATTTATATATGTCGAAGATTTCTTGATTAATATGAT 4260
Qy 4261 AATAGATTTATTAACCAACATATATCTAATTTAGAGTTATCTCGTAAATGATGAGAT 4320
Db 4261 AATAGATTTATTAACCAACATATATCTAATTTAGAGTTATCTCGTAAATGATGAGAT 4320
Qy 4321 ACTGATGATTTCTGAGAAATTTTGGCTTAACTATATGAGAAACGATGGAAGATTTAAATAT 4380
Db 4321 ACTGATGATTTCTGAGAAATTTTGGCTTAACTATATGAGAAACGATGGAAGATTTAAATAT 4380
Qy 4381 TACAGAGAAATTTAATGCGGCTTACGAGATGCTGAAATTTGCGACTAGAACATGACACA 4440
Db 4381 TACAGAGAAATTTAATGCGGCTTACGAGATGCTGAAATTTGCGACTAGAACATGACACA 4440
Qy 4441 GGTAAATATATGCTTTTGTGCGATTTCTGAGTCTATATAGAAATTTGCAATTTGCGAGAG 4500
Db 4441 GGTAAATATATGCTTTTGTGCGATTTCTGAGTCTATATAGAAATTTGCAATTTGCGAGAG 4500
Qy 4501 ATGATGATATATATATATGAGTATATATGCGATATATGAGAGATAGATTTTGTGTTAGTA 4560
Db 4501 ATGATGATATATATATATGAGTATATATGCGATATATGAGAGATAGATTTTGTGTTAGTA 4560
Qy 4561 GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAATTTTCATGCTTTAACAGAGAA 4620
Db 4561 GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAATTTTCATGCTTTAACAGAGAA 4620
Qy 4621 GAGCTGTAAAGAAATTTTGTGCGAGATCTATATAGAAAAATATGTTTGGTGAAGCTT 4680
Db 4621 GAGCTGTAAAGAAATTTTGTGCGAGATCTATATAGAAAAATATGTTTGGTGAAGCTT 4680
Qy 4681 TATTCACGAGATATTAATAAGATATTAATTTCCAAATTAATATAGAAATGATTTGATGAG 4740
Db 4681 TATTCACGAGATATTAATAAGATATTAATTTCCAAATTAATATAGAAATGATTTGATGAG 4740
Qy 4741 GATTTGCTTTTAAATTTGAGAGTCTTGAAACAATGTAACGCTGATATGTTGATACTAGA 4800
Db 4741 GATTTGCTTTTAAATTTGAGAGTCTTGAAACAATGTAACGCTGATATGTTGATACTAGA 4800

Qy 4801 GAATATATTTATATATATATGATCTTGTATACAGTTCGCTTATTAATCAGAAAAATTTCTATAT 4860
Db 4801 GAATATATTTATATATATATATGATCTTGTATACAGTTCGCTTATTAATCAGAAAAATTTCTATAT 4860
Qy 4861 AATATATTTGATTTGTGACAAAGATTTGGAGATTAACCCCTTTAATAGTTAAAAAGAGTTT 4920
Db 4861 AATATATTTGATTTGTGACAAAGATTTGGAGATTAACCCCTTTAATAGTTAAAAAGAGTTT 4920
Qy 4921 AGTCATTTATTTTGTAGCAAAAAGTTATTAAGAAAGTTAAATGTTTAAACAAAATGATAT 4980
Db 4921 AGTCATTTATTTTGTAGCAAAAAGTTATTAAGAAAGTTAAATGTTTAAACAAAATGATAT 4980
Qy 4981 TCAACAGATTTGTTGATATAGATTTCTTGCAATTTATAGATCTTATTCGAAAAAGAAATA 5040
Db 4981 TCAACAGATTTGTTGATATAGATTTCTTGCAATTTATAGATCTTATTCGAAAAAGAAATA 5040
Qy 5041 CGTAAATATCCATTTATTAACGGAAGATATTTATCAAGAAACATTTAGTTACGTTG 5100
Db 5041 CGTAAATATCCATTTATTAACGGAAGATATTTATCAAGAAACATTTAGTTACGTTG 5100
Qy 5101 TATTTGATGAATTTTTCGCTTAACCTATATGTAATGTTATATTAAGAAATTTTCAAAAGCAG 5160
Db 5101 TATTTGATGAATTTTTCGCTTAACCTATATGTAATGTTATATTAAGAAATTTTCAAAAGCAG 5160
Qy 5161 TAGAGTAAAAATGATAAAAATTTAGTATTGTTCCAGTTATTAATAGTAAATATAT 5220
Db 5161 TAGAGTAAAAATGATAAAAATTTAGTATTGTTCCAGTTATTAATAGTAAATATAT 5220
Qy 5221 TAAGATGTTGTAAGAAACATTTATTAACAAAATTAATAAAATATAGAAATATATTTGA 5280
Db 5221 TAAGATGTTGTAAGAAACATTTATTAACAAAATTAATAAAATATAGAAATATATTTGA 5280
Qy 5281 TAGATGATGCTCTGATAGATGATTCGCTAAATATGCAAGAAATATGCAAAAAAGATA 5340
Db 5281 TAGATGATGCTCTGATAGATGATTCGCTAAATATGCAAGAAATATGCAAAAAAGATA 5340
Qy 5341 AAAAGTAAAAATTTTTCCTACTATCATATGAGAGATATCAATGCTAGAAATCATAGAA 5400
Db 5341 AAAAGTAAAAATTTTTCCTACTATCATATGAGAGATATCAATGCTAGAAATCATAGAA 5400
Qy 5401 TAAAGCGGAGTACAGCTGAATATATATGTTTGTGACCTGATGATGTTGTTGATAGTA 5460
Db 5401 TAAAGCGGAGTACAGCTGAATATATATGTTTGTGACCTGATGATGTTGTTGATAGTA 5460
Qy 5461 GATTTAGTAAAAATTTATATTTTAAATATTAATAAAGTAAAGTAAATATCTGTTGTT 5520
Db 5461 GATTTAGTAAAAATTTATATTTTAAATATTAATAAAGTAAAGTAAATATCTGTTGTT 5520
Qy 5521 TGTAGCTACTTTTTCAGAAAAATATATATTTTGAAGTAAATTCAAAAATATGATAT 5580
Db 5521 TGTAGCTACTTTTTCAGAAAAATATATATTTTGAAGTAAATTCAAAAATATGATAT 5580
Qy 5581 TTGAAGCAATTAATACCGTGCAGAGACATGAGGAAAAAATTTTATGAATTTGATATATA 5640
Db 5581 TTGAAGCAATTAATACCGTGCAGAGACATGAGGAAAAAATTTTATGAATTTGATATATA 5640
Qy 5641 ATATATTTTTCCTACTCTGTTGTTGTAACCTATATTAAGAAAGATATCATACAGATCTTT 5700
Db 5641 ATATATTTTTCCTACTCTGTTGTTGTAACCTATATTAAGAAAGATATCATACAGATCTTT 5700
Qy 5701 TTCAAGAGATCAATGTTTGAAGAAAGATTTTACTTTTAAATTCGATTTATTTAAAGATAT 5760
Db 5701 TTCAAGAGATCAATGTTTGAAGAAAGATTTTACTTTTAAATTCGATTTATTTAAAGATAT 5760
Qy 5761 TAGATGAGATTTAGTTATTTGACAGCAATCTTTATTTTATAGAGAGATATCTAAGTA 5820
Db 5761 TAGATGAGATTTAGTTATTTGACAGCAATCTTTATTTTATAGAGAGATATCTAAGTA 5820
Qy 5821 CAGTAAATCTTTTAAAGAAAGTGTGTTTTCGAATTTGCAAAATTTGCAAAACAAAGTGA 5880
Db 5821 CAGTAAATCTTTTAAAGAAAGTGTGTTTTCGAATTTGCAAAATTTGCAAAACAAAGTGA 5880

QY 5881 TAGATATGTTAAAGCAATATATGTGAGGATTTTGAAGTATCAATGTTAAAGATCTA 5940
 DB 5881 TAGATATGTTAAAGCAATATATGTGAGGATTTTGAAGTATCAATGTTAAAGATCTA 5940
 QY 5941 TAGCTGGCAAGTATTTTATTTATGCTTACTAATGTTAAATACGGAACACGCTATTT 6000
 DB 5941 TAGCTGGCAAGTATTTTATTTATGCTTACTAATGTTAAATACGGAACACGCTATTT 6000
 QY 6001 TTGACAAATTTTAAATTTTAAAGATCTTTATTAATAAATATTTATTTAACTGTTAAAG 6060
 DB 6001 TTGACAAATTTTAAATTTTAAAGATCTTTATTAATAAATATTTATTTAACTGTTAAAG 6060
 QY 6061 TATCTAACAAATTTCTTGTCTTAAATTTTGTATAGAAATGTTTCCAGCAAGTTT 6120
 DB 6061 TATCTAACAAATTTCTTGTCTTAAATTTTGTATAGAAATGTTTCCAGCAAGTTT 6120
 QY 6121 TTAATAAATATTTATGTTTAAATAGAGATATCATGATACTATTAGTAAATTTTCTA 6180
 DB 6121 TTAATAAATATTTATGTTTAAATAGAGATATCATGATACTATTAGTAAATTTTCTA 6180
 QY 6181 TAAATGACTAT 6240
 DB 6181 TAAATGACTAT 6240
 QY 6241 ATCAGACCTACAAACATATATAGATTTCTTCTGTGATGACGGATGATGATATTCGG 6300
 DB 6241 ATCAGACCTACAAACATATATAGATTTCTTCTGTGATGACGGATGATGATATTCGG 6300
 QY 6301 AAGAAATTTGTTTATGATATGCGAAGAAATATGATGATGATGATGATGATGATGATG 6360
 DB 6301 AAGAAATTTGTTTATGATATGCGAAGAAATATGATGATGATGATGATGATGATGATG 6360
 QY 6361 ACGCGCGGCTATCATGATGCGGCTATATATGCGATATGCGGCGCAAGGCTACTTAG 6420
 DB 6361 ACGCGCGGCTATCATGATGCGGCTATATATGCGATATGCGGCGCAAGGCTACTTAG 6420
 QY 6421 CTATTATGACTAGATGATTTTATTCATTCGAGGATTCATCCAAAGTTTACAGAAACAA 6480
 DB 6421 CTATTATGACTAGATGATTTTATTCATTCGAGGATTCATCCAAAGTTTACAGAAACAA 6480
 QY 6481 TTGAGAGAGAAATGCGCTTGTGCGAGTTGCTGTTATGATAGGATGATGCTTGGGGC 6540
 DB 6481 TTGAGAGAGAAATGCGCTTGTGCGAGTTGCTGTTATGATAGGATGATGCTTGGGGC 6540
 QY 6541 ATTTCTTAACAGAGAGCGGCTTCTTCAATACAGGCTGTTCTGAGCGGCGAGATGTT 6600
 DB 6541 ATTTCTTAACAGAGAGCGGCTTCTTCAATACAGGCTGTTCTGAGCGGCGAGATGTT 6600
 QY 6601 GTAAAAAGCTGTAAGAGCGGATGTCATGCTTGTGAGGCTGTAATAACTCTATA 6660
 DB 6601 GTAAAAAGCTGTAAGAGCGGATGTCATGCTTGTGAGGCTGTAATAACTCTATA 6660
 QY 6661 AAAAAGAACTATTTGAAGATTTTGAATTTGAAAAAGGTAAGATTCATGAAGATGATAT 6720
 DB 6661 AAAAAGAACTATTTGAAGATTTTGAATTTGAAAAAGGTAAGATTCATGAAGATGATAT 6720
 QY 6721 TCACTATGCTGCTCTATAGATTAAGAAAAAGTTGCAATAGTTAAGAGAGTCTTGTACT 6780
 DB 6721 TCACTATGCTGCTCTATAGATTAAGAAAAAGTTGCAATAGTTAAGAGAGTCTTGTACT 6780
 QY 6781 ATATATGTAAGCAAGAAATATATATCAACTCTGACATGACATGACATGACATGACAT 6840
 DB 6781 ATATATGTAAGCAAGAAATATATATCAACTCTGACATGACATGACATGACATGACAT 6840
 QY 6841 GCGTACTGAAATTTCAAAATGAAGATGACATTTCTATGAAAGTAAAGATTAAGAGC 6900
 DB 6841 GCGTACTGAAATTTCAAAATGAAGATGACATTTCTATGAAAGTAAAGATTAAGAGC 6900
 QY 6901 TCTTACTAGAGTGTATGCTTCAATTTTACGCTTTGCTTTTGTATTTAGGCAAAATATA 6960
 DB 6901 TCTTACTAGAGTGTATGCTTCAATTTTACGCTTTGCTTTTGTATTTAGGCAAAATATA 6960
 QY 6961 ATCATTGTTGAGCAACACCAAAAGAAAGCTT 6992

DB 6961 ATCATTGTTGAGCAACACCAAAAGAAAGCTT 6992
 RESULT 2
 AAZ60929
 ID AAZ60929 standard; DNA; 26281 BP.
 XX
 AC AAZ60929;
 XX
 DT 30-MAY-2000 (first entry)
 XX
 DE Nucleotide sequence of a capsular gene cluster of *S. suis* serotype 2.
 KW Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
 KW capsular component; antigen; regulation; chain length determination;
 KW complement-mediated opsonophagocytosis; serotype-specific detection;
 KW antigen; vaccine; Streptococcal disease; ORF 2Z; ORF 2Y; ORF 2Z; Cps2A;
 KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;
 KW Cps2O; Cps2P; Cps2R; Cps2S; Cps2T; ss.
 XX
 OS Streptococcus suis.
 XX
 FH Key
 FH 2. .721
 FT CDS
 FT /tag= a
 FT /note= "ORF 2Z; encodes AAY68950"
 FT complement (822. .2079)
 FT CDS
 FT /tag= b
 FT /note= "ORF 2Y; encodes AAY68951"
 FT 2202. .2936
 FT CDS
 FT /tag= c
 FT /transl_except= (pos: 2916. .2918, aa: Xaa)
 FT /note= "ORF 2X; Xaa is an unspecified amino acid; encodes
 FT AAY68952"
 FT CDS
 FT 3041. .4486
 FT /tag= d
 FT /note= "Cps2A; encodes AAY68953"
 FT 4504. .5193
 FT CDS
 FT /tag= e
 FT /note= "Cps2B; encodes AAY68954"
 FT 5203. .5880
 FT /tag= f
 FT /note= "Cps2C; encodes AAY68955"
 FT 5919. .6650
 FT CDS
 FT /tag= g
 FT /note= "Cps2D; encodes AAY68956"
 FT 6675. .8054
 FT CDS
 FT /tag= h
 FT /note= "Cps2E; encodes AAY68957"
 FT 8089. .9258
 FT CDS
 FT /tag= i
 FT /note= "Cps2F; encodes AAY68958"
 FT 9262. .10419
 FT CDS
 FT /tag= j
 FT /transl_except= (pos: 10057. .10059, aa: Xaa)
 FT /note= "Cps2G; Xaa is an unspecified amino acid; encodes
 FT AAY68959"
 FT CDS
 FT 10808. .12176
 FT /tag= k
 FT /transl_except= (pos: 11963. .11965, aa: Xaa)
 FT /note= "Cps2H; Xaa is an unspecified amino acid; encodes
 FT AAY68960"
 FT CDS
 FT 12213. .13445
 FT /tag= l
 FT /note= "Cps2I; encodes AAY68961"
 FT 13583. .14581
 FT CDS
 FT /tag= m
 FT /note= "Cps2J; encodes AAY68962"
 FT 14574. .15578
 FT CDS
 FT /tag= n
 FT /note= "Cps2K; encodes AAY68963"
 FT 18401. .18904
 FT CDS


```
FT      /tag= 0
FT      /transl_except= (pos: 18755..18957, aa: Ile)
FT      /transl_except= (pos: 18770..18772, aa: Leu)
FT      /transl_except= (pos: 18772..18778, aa: Ile)
FT      /transl_except= (pos: 18778..18778, aa: Ile)
FT      /transl_except= (pos: 18836..18838, aa: Pro)
FT      /transl_except= (pos: 18890..18892, aa: Thr)
FT      /transl_except= (pos: 18898..18898, aa: Pro)
FT      /transl_except= (pos: 18935..18937, aa: Pro)
FT      /transl_except= (pos: 18953..18955, aa: Leu)
FT      /transl_except= (pos: 18955..18970, aa: Xaa)
FT      /transl_except= (pos: 18971..18973, aa: Trp)
FT      /transl_except= (pos: 18977..18979, aa: Ser)
FT      /transl_except= (pos: 18980..18982, aa: Gln)
FT      /transl_except= (pos: 18989..18991, aa: Asn)
FT      /note= "Cps20; Xaa is an unspecified amino acid; encodes
FT      AAY68964"
FT      20327..21343
FT      CDS
FT      /tag= P
FT      /note= "Cps2P; encodes AAY68965"
FT      21355..21867
FT      CDS
FT      /tag= Q
FT      /note= "Cps2Q; encodes AAY68966"
FT      21933..22487
FT      CDS
FT      /tag= R
FT      /note= "Cps2R; encodes AAY68967"
FT      22501..23127
FT      CDS
FT      /tag= S
FT      /note= "Cps2S; encodes AAY68968"
FT      23136..24368
FT      CDS
FT      /tag= T
FT      /note= "Cps2T; encodes AAY68969"
FT      24368..25000
FT      03-FEB-2000.
FT      19-JUL-1999; 99MO-NL000460.
FT      22-JUL-1998; 98RP-00202465.
FT      22-JUL-1998; 98RP-00202467.
FT      (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
FT      Smith HE;
FT      WPI; 2000-195104/17.
FT      P-PSDB; AAY68950, AAY68951, AAY68952, AAY68953, AAY68954, AAY68955,
DR      AAY68956, AAY68957, AAY68958, AAY68959, AAY68960, AAY68961, AAY68962,
DR      AAY68963, AAY68964, AAY68965, AAY68966, AAY68967, AAY68968, AAY68969.
XX      New nucleic acid containing the capsular gene cluster of Streptococcus
PT      suis, used for serotype-specific detection and to generate antigens or
PT      mutants for vaccination.
XX      Claim 4; Fig 3; 144P; English.
XX      The present sequence represents the capsular gene cluster of
XX      Streptococcus suis serotype 2. The genes in this cluster are involved in
XX      polysaccharide biosynthesis of capsular components and antigens. The
XX      proteins are involved in regulation (CpsA), chain length determination
XX      (CpsB, CpsC), export (CpsD), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
XX      CpsJ, CpsK). The capsule confers bacterium resistance to complement-
XX      mediated opsonophagocytosis. The gene cluster is used as a source of
XX      probes and primers for serotype-specific detection of S. suis and is also
XX      useful for recombinant production of the proteins. The proteins are then
XX      useful for producing antigens that can be used in vaccines, for
XX      controlling or eradicating a Streptococcal disease, in humans or animals,
XX      e.g. against S. suis in pigs
XX      Sequence 26281 BP; 8252 A; 4158 C; 5488 G; 8383 T; 0 U; 0 Other;
SQ
```

Query Match 8.7%; Score 611; DB 3; Length 26281;
Best Local Similarity 99.6%; Pred. No. 2.1e-248;

```
Matches 761; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 6229 ATAGCATTTGAATCAGACCTTACAAACATATAGAGATTTCTGTGTAATGACGGTAGTA 6288
DB 14638 ATAGCATTTGAATCAGACCTTACAAACATATAGAGATTTCTGTGTAATGACGGTAGTA 14697
QY 6289 CGGATTAATTCGGAAGAAATTTGTTTATGATATAGCGAAGAAATGTCGATTCGTTAT 6348
DB 14698 CGGATTAATTCGGAAGAAATTTGTTTATGATATAGCGAAGAAATGTCGATTCGTTAT 14757
QY 6349 TTTAAAAAGAGAACGGCGGCTATCAGATGCCGTAATTATGATCGCGCAAG 6408
DB 14758 TTTAAAAAGAGAACGGCGGCTATCAGATGCCGTAATTATGATCGCGCAAG 14817
QY 6409 GTGACTACTTACCTTTATATAGCTAGATGATTTATTCATTCGAGATTCACCGTT 6468
DB 14818 GTGACTACTTACCTTTATATAGCTAGATGATTTATTCATTCGAGATTCACCGTT 14877
QY 6469 TACACGAGCAATTTGAGAGAGAAATGCCCTTGTGCACTTCTGTTATGATAGGTAG 6528
DB 14878 TACACGAGCAATTTGAGAGAGAAATGCCCTTGTGCACTTCTGTTATGATAGGTAG 14937
QY 6529 ATGCTTCGGGGGATTTCTTAACAGCAGACCGCTTCTTAATCAGCTGTTCTGAGCG 6588
DB 14938 ATGCTTCGGGGGATTTCTTAACAGCAGACCGCTTCTTAATCAGCTGTTCTGAGCG 14997
QY 6589 GCAGAAATGTTTGTAAAAAGCTCTAGAGCGGAGATGTCATGCTTTGTGTGCGCTGTA 6648
DB 14998 GCAGAAATGTTTGTAAAAAGCTCTAGAGCGGAGATGTCATGCTTTGTGTGCGCTGTA 15057
QY 6649 ATTAATCTTATAAAAAGACTTATTTGAAAGATTTTGGATTTTGAAGGGTATGATTCAG 6708
DB 15058 ATTAATCTTATAAAAAGACTTATTTGAAAGATTTTGGATTTTGAAGGGTATGATTCAG 15117
QY 6709 AAGATGAATACTTCACTTATGCTTATGCTCTATGAGTTAGAAAAAGTGCATATGATTAAG 6768
DB 15118 AAGATGAATACTTCACTTATGCTTATGCTCTATGAGTTAGAAAAAGTGCATATGATTAAG 15177
QY 6769 AGTCTTGTACTTATTTATGTTGACCGAATAATAGTATCAAACTTGTAGATGATGACC 6828
DB 15178 AGTCTTGTACTTATTTATGTTGACCGAATAATAGTATCAAACTTGTAGATGATGACC 15237
QY 6829 ATGCTTCCATTGCTTACTGGAATTTCAAAATGAAAGAAATGATGAAAGTAAAG 6888
DB 15238 ATGCTTCCATTGCTTACTGGAATTTCAAAATGAAAGAAATGATGAAAGTAAAG 15297
QY 6889 GAGATTAAGAGCTTACTAGAGTGTATGCTTATGCTTATGCTTATGCTTATGCTTAT 6948
DB 15298 GAGATTAAGAGCTTACTAGAGTGTATGCTTATGCTTATGCTTATGCTTATGCTTAT 15357
QY 6949 TAGGCAAAATTAATCATTTGTTGAGCAAAACGCAAAAGAGCTT 6992
DB 15358 TAGGCAAAATTAATCATTTGTTGAGCAAAACGCAAAAGAGCTT 15401
RESULT 3
ID ABA01441 standard; DNA; 17468 BP.
XX ABA01441;
XX 21-FEB-2002 (first entry)
XX Streptococcus thermophilus eps3 operon #2.
XX Streptococcus thermophilus eps3 operon #2.
XX Ekopolysaccharide; lactic acid bacterium; eps; fermented food product;
XX ds.
XX Streptococcus thermophilus.
XX OS
XX WO200179500-A2.
XX 25-OCT-2001.
XX PD
```

```

XX PF 18-APR-2001; 2001MO--PR001199.
XX PR 18-APR-2000; 2000FR-00004972.
XX XX (INRG ) INRA INST RECH AGRONOMIQUE.
PA (DANO-) CTE DANONE SA GERAVALS.
PA (RHOD ) RHODIA CHIM.
XX PI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX DR WPI; 2002-017616/02.
XX XX
XX PT New nucleic acid fragments containing exopolysaccharide operon, useful
XX FT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX PS Claim 9; Page 75-80; 144pp; French.
XX XX
CC The present sequence is an eps operon from Streptococcus thermophilus.
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC synthesis. The operon is useful for producing chimeric eps operons, for
CC optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC mouth feel and rheological properties to fermented food products (e.g.
CC yoghurt). They function as thickeners, to provide free-flowing and creamy
CC texture, and may also have biological activities beneficial to health
XX SO Sequence 17468 BP; 5871 A; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;
XX
Query Match 0.4%; Score 28; DB 6; Length 17468;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1818 TTTTATGATTTTGTAAACAGTAGAGAC 1845
DB 8126 TTTTATGATTTTGTAAACAGTAGAGAC 8153
RESULT 4
ADSI13200
ID ADSI13200 standard; DNA; 16032 BP.
XX AC ADSI13200;
XX DT 16-DEC-2004 (first entry)
XX DE S. thermophilus CNCM I-2980 polysaccharide synthesis-related operon DNA.
XX XX lactic acid bacterium; food; meat; cereal; dairy;
XX KM polysaccharide synthesis; operon; ds.
XX OS Streptococcus thermophilus.
XX XX
XX FH Location/Qualifiers
XX FT 342..1802
XX FT /*tag= a
XX FT /product= "Transcriptional regulator eps13A protein"
XX FT /note= "The corresponding protein sequence is not shown
XX FT shown within the specification"
XX FT 1803..2534
XX FT /*tag= b
XX FT /product= "Polysaccharide polymerisation/export-related
XX FT eps13B protein"
XX FT /note= "The corresponding protein sequence is not shown
XX FT shown within the specification. The CDS has a prokaryotic
XX FT val stat."
XX FT 2543..3235
XX FT /*tag= c
XX FT /product= "Polysaccharide polymerisation/export-related
XX FT eps13C protein"
XX FT /note= "The corresponding protein sequence is not shown
XX FT shown within the specification"
XX FT 3245..3585
XX FT /*tag= d
XX FT CDS

```

PT	/product= "Polysaccharide polymerisation/export-related
PT	eps13J protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	4042. .5409
CDS	/tag= e
PT	/product= "Undecaprenyl-phosphate glycosyltransferase
PT	eps13E protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification. The CDS has a prokaryotic
PT	Val start."
PT	5611. .6195
CDS	/tag= f
PT	/product= "Undecaprenyl-phosphate glycosyltransferase
PT	eps13F protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification. The CDS has a prokaryotic
PT	Val start."
PT	6251. .6634
CDS	/tag= g
PT	/product= "Undecaprenyl-phosphate glycosyltransferase
PT	eps13G protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	6643. .7092
CDS	/tag= h
PT	/product= "Beta-1,4-galactosyltransferase eps13H protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	7092. .7607
CDS	/tag= i
PT	/product= "Beta-1,4-galactosyltransferase eps13I protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	7597. .8493
CDS	/tag= j
PT	/product= "Rhamnosyltransferase eps13J protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification. The CDS has a prokaryotic
PT	Val start."
PT	8763. .9797
CDS	/tag= k
PT	/product= "Glycosyltransferase eps13K protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	9827. .10969
CDS	/tag= l
PT	/product= "Repetitive unit polymerase eps13L protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	10984. .11793
CDS	/tag= m
PT	/product= "Repetitive unit polymerase eps13M protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	11844. .12578
CDS	/tag= n
PT	/partial
PT	/product= "Glycosyltransferase eps13N protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification. No start codon."
PT	12633. .13016
CDS	/tag= o
PT	/product= "Glycosyltransferase eps13O protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification"
PT	13049. .14482
CDS	/tag= p
PT	/product= "Transmembrane transporter eps13P protein"
PT	/note= "The corresponding protein sequence is not shown
PT	shown within the specification. The CDS has a prokaryotic
PT	Val start."
PT	complement(14614. .15870)

FT	/*tag= q	
FT	/product= "Transposase IS1193 protein"	
FT	/note= "The corresponding protein sequence is not shown	
FT	shown within the specification"	
XX		
PN	FR2852604-A1.	
XX		
PD	24-SEP-2004.	
XX		
PF	17-MAR-2003; 2003FR-00003242.	
XX		
PR	17-MAR-2003; 2003FR-00003242.	
XX		
PA	(RHOD) RHODIA CHIM SAS.	
XX		
PI	Horvath P, Manoury E, Huppert S, Fremaux C;	
DR	WPI; 2004-671092/66.	
XX		
XX	New strain of lactic acid bacteria, useful for preparation of foods and	
PT	pharmaceuticals, especially fermented dairy products, contains genes	
PT	involved in polysaccharide biosynthesis.	
XX		
PS	Claim 5; SEQ ID NO 1; 35pp; French.	
XX		
CC	The invention relates to a novel strain of a lactic acid bacterium that	
CC	contains at least one of 8 specific nucleic acid sequences all of which	
CC	are defined in the specification. The bacterium of the invention may be	
CC	useful during food preparation, particularly that of beverages and meat,	
CC	cereal, or dairy products e.g. fermented milk, yoghurt, 'matured' cream,	
CC	dessert cream, cheese, soft cheeses, cheese spreads, cottage cheese, milk-	
CC	based drinks, dairy product recheetes and baby milk, where the products	
CC	are derived from an animal and/or plant. The current sequence is that of	
CC	the Streptococcus thermophilus CNCM I-2380 polysaccharide synthesis-	
CC	related operon DNA of the invention.	
XX		
SO	Sequence 16032 BP; 5191 A; 2323 C; 3188 G; 5330 T; 0 U; 0 Other;	
XX		
Query Match	0.4%; Score 27; DB 13; Length 16032;	
Best Local Similarity	100.0%; Pred. No. 0.69;	
Matches 27; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	1807 TGGGGAGTATTTTAAATGATTTTGT 1833	
DB	7076 TGGGGAGTATTTTAAATGATTTTGT 7102	
XX		
RESULT 5		
ABN67430		
ID	ABN67430 standard; DNA; 447 BP.	
XX		
AC	ABN67430;	
XX		
DT	01-JUL-2002 (first entry)	
XX		
DE	Streptococcus polynucleotide SEQ ID NO 2773.	
XX		
KW	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;	
KW	group A streptococcus; Streptococcus pyogenes; antibacterial; gene;	
XX	antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.	
XX		
OS	Streptococcus agalactiae.	
XX		
PN	WO200234771-A2.	
XX		
PD	02-MAY-2002.	
XX		
PF	29-OCT-2001; 2001WO-GB004789.	
XX		
PR	27-OCT-2000; 2000GB-00026333.	
PR	24-NOV-2000; 2000GB-00028727.	
PR	07-MAR-2001; 2001GB-00005640.	
XX		

PA	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
PI	Telford J, Massignani V, Margarit Y Rosi, Grandi G, Frazer C;
PI	Tetteijn H;
DR	WPI: 2002-352536/38.
DR	P-PSDB: ABP26799.
XX	
XX	New Streptococcus protein for the treatment or prevention of infection or
PT	disease caused by Streptococcus bacteria, such as meningitis, and for
PT	detecting a compound that binds to the protein.
XX	
PS	Claim 7; Page 3428; 4525pp: English.
XX	
CC	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC	the specification. The proteins have antibacterial and anti-inflammatory
CC	activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC	antibodies that bind (I) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins
XX	
SQ	Sequence 447 BP; 150 A; 59 C; 93 G; 145 T; 0 U; 0 Other;
XX	
Query Match	0.4%; Score 26; DB 6; Length 447;
Beet Local Similarity	100.0%; Pred. No. 2.3;
Matches	26; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY	1755 TTTATTGTTTCAGTGGAGGAATGAA 1780
DB	382 TTTATTGTTTCAGTGGAGGAATGAA 407
RESULT 6	
ADV84473	
ID	ADV84473 standard; DNA; 450 BP.
XX	
AC	ADV84473;
XX	
XX	24-FEB-2005 (first entry)
XX	
DE	Streptococcus agalactiae DNA sequence, SEQ ID 5614.
XX	
KM	Antibacterial; vaccine; bacterial infection; ds.
XX	
OS	Streptococcus agalactiae.
XX	
PN	WO200292818-A2.
XX	
PD	21-NOV-2002.
XX	
PF	26-APR-2002; 2002WO-IB003059.
XX	
PR	26-APR-2001; 2001FR-00005642.
XX	
PA	(INSP) INST PASTEUR.
XX	
PA	(CNRS) CNRS CENT NAT RECH SCTI.
XX	
PI	Glaeser P, Ruenick C, Chevalier F, Frangoul L, Laloui L;
PI	Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
DR	WPI: 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 4; SEQ ID NO 5614; 439bp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons, biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (i) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (i) or (ii) are
CC useful for treatment of a bacterial S. agalactiae infection. The complete
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
CC contains only 2344 sequences.
XX
SQ Sequence 450 BP; 153 A; 58 C; 92 G; 147 T; 0 U; 0 Other;
XX
Query Match 0.4%; Score 26; DB 13; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
Db 382 TTTATTGTTCAAGTGGAGAAATGAA 407
XX
RESULT 7
ADM79774
ID ADM79774 standard; DNA; 2226 BP.
XX
AC ADM79774;
XX
DT 03-JUN-2004 (first entry)
XX
DE Group B Streptococcus cpsd-cdsr-cpsf-cpsg partial consensus DNA sequence.
XX
KM group B streptococcus; GBS bacterium; cpsd; cpsr; cpsf; cpsg; cpsl gene;
XX M gene; GBS infection; ds.
XX
OS Streptococcus sp. 'group B'.
XX
PN WO2003025216-A1.
XX
PD 27-MAR-2003.
XX
PF 18-SEP-2002; 2002WO-AU01281.
XX
PR 19-SEP-2001; 2001AU-00007749.
XX
PA (MSYD-) WESTERN SYDNEY AREA HEALTH SERVICE.
XX
PI Panrong K, Gilbert G;
XX
DR WPI; 2003-381495/36.
XX
XX Typing a group B streptococcus (GBS) bacterium for diagnosing the GBS
PT infections in pregnant women, elderly or immunocompromised patients by
PT analyzing the sequence of the regions in the cpsd, cpsr, cpsf, cpsg or
PT cpsl/M gene of the bacterium.
XX
PS Claim 2; Fig 1; 106pp; English.
XX
CC This invention relates to a novel method of typing a group B

CC streptococcus (GBS) bacterium which comprises analyzing the nucleotide
CC sequence of one or more regions within the cpsd, cpsr, cpsf, cpsg and/or
CC cpsl/M genes of the bacterium, where the regions comprise one or more
CC nucleotides having sequences that vary between types. The method is
CC useful for preparing a composition for serotyping and/or subtyping a GBS
CC bacterium for diagnosing GBS infections in pregnant women, elderly and/or
CC immunocompromised patients. The present sequence is that of a consensus
CC DNA sequence for the 3' end of the cpsd-cpsr-cpsf and the 5' end of the
CC cpsg sequences of group B Streptococci which is related to the method of
CC the invention.
XX
SQ Sequence 2226 BP; 767 A; 293 C; 417 G; 749 T; 0 U; 0 Other;
XX
Query Match 0.4%; Score 26; DB 11; Length 2226;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
Db 1941 TTTATTGTTCAAGTGGAGAAATGAA 1966
XX
RESULT 8
ABK90550
ID ABK90550 standard; DNA; 6865 BP.
XX
AC ABK90550;
XX
DT 15-NOV-2002 (first entry)
XX
DE Betal,3-galactose transferase DNA #2.
XX
KM Betal,3-galactose transferase; gene; ds; galactose; N-acetylglucosamine;
XX betal,3-bond; galactose-containing glucide; uridine-5'-diphosphate;
XX infection.
XX
OS Streptococcus agalactiae.
XX
FH Key Location/Qualifiers
FT CDS 617..1792
FT /tag= a
FT /product= "Betal,3-galactose transferase #2"
FT 1816..2262
FT /tag= b
FT /product= "Betal,3-galactose transferase #3"
FT /partial
FT /note= "No stop codon given"
FT CDS 2265..2747
FT /tag= c
FT /product= "Betal,3-galactose transferase #4"
FT 2843..3879
FT /tag= d
FT /product= "Betal,3-galactose transferase #5"
FT /partial
FT /note= "No stop codon given"
FT CDS 3982..4956
FT /tag= e
FT /product= "Betal,3-galactose transferase #6"
FT 5009..5950
FT /tag= f
FT /product= "Betal,3-galactose transferase #7"
XX
PN JP2002199885-A.
XX
PD 16-JUL-2002.
XX
PF 05-JAN-2001; 2001JP-00000392.
XX
PR 05-JAN-2001; 2001JP-00000392.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR WPI; 2002-612563/66.
XX

DR P-PSDB; ABG31703, ABG31704, ABG31705, ABG31706, ABG31707, ABG31708.
XX
PT Beta1,3-galactose transferase and a DNA encoding the enzyme, a protein
PT having beta1,3-galactose transferase activity, a transformant, a protein
PT preparation of a protein, and preparation of a galactose-containing
PT glucide.
XX
PS Example 1; Page 15-22; 24pp; Japanese.
XX
CC The invention relates to a protein having beta1,3-galactose transferase
CC activity, derived from a microbe having an activity of transferring
CC galactose to N-acetylglucosamine by beta1,3-bond. The protein can be used
CC in a method for the preparation of a protein having beta1,3-galactose
CC transferase activity in which the above transformant is cultured in a
CC medium to form and accumulate a protein having beta1,3-galactose
CC transferase activity, and in a method for the preparation of a galactose-
CC containing glucide in which a culture liquid of the above transformant or
CC a treated product of the culture liquid is used as an enzyme source, the
CC enzyme source uridine-5'-diphosphate galactose and a receptor glucide are
CC made to be present in an aqueous medium, and a galactose-containing
CC glucide is formed and accumulated in the aqueous medium. The galactose-
CC containing glucide is used as a candidate for infection preventors. This
CC sequence represents DNA encoding beta1,3-galactose transferase proteins
CC of the invention
XX
SQ Sequence 6665 BP; 2453 A; 847 C; 1185 G; 2380 T; 0 U; 0 Other;
Query Match 0.4%; Score 26; DB 6; Length 6665;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
|||||
Db 2197 TTTATTGTTCAAGTGGAGAAATGAA 2222
|||||
RESULT 9
ACA64723
ID ACA64723 standard; DNA: 17276 BP.
XX
AC ACA64723;
XX
DT 18-JUN-2003 (first entry)
XX
DE Streptococcus capsular polysaccharide gene.
XX
XX Superantigen; ds; gene; SAg; staphylococcal enterotoxin; tumour; cancer;
XX apoptosis; gene therapy; mammalian cell receptor; cytosstatic;
XX tumour associated lipid; anergy; T cell; antigen presenting cell; APC;
XX tumouricidal immunocyte; antitumour.
XX
OS Streptococcus sp.
XX
PN US2002177551-A1.
XX
PD 28-NOV-2002.
XX
PF 30-MAY-2001; 2001US-00870759.
XX
PR 31-MAY-2000; 2000US-0208128P.
XX
PA (TERM/) TERMAN D S.
XX
PI Terman DS;
XX
DR WPI; 2003-361759/34.
XX
DR P-PSDB; ABU79115.
XX
PT A mammalian cell receptor, useful in the treatment of cancer by binding
PT to tumor associated lipids where the binding induces anergy or apoptosis
PT in T cells and antigen presenting cells.
XX
PS Example 2; Page; 167pp; English.

XX
CC The invention relates to a mammalian cell receptor, useful in the
CC treatment of cancer, which binds to tumour associated lipids and induces
CC anergy or apoptosis in the T cells and antigen presenting cells (APCs).
CC Also included are a mammalian cell useful in the treatment of cancer
CC where the receptor which binds tumour associated lipids and induces
CC cellular inactivation or death is deleted or functionally deactivated,
CC producing (M1) a tumouricidal immunocyte population in vivo in a mammal
CC (by allowing tumour associated lipids to contact immunocytes in which
CC receptors for immunosuppressive fatty acids, ceramides, glycolipids,
CC sphingolipids, glycosphingolipids, phosphosphingolipids, gangliosides,
CC sialylated glycans, lipopeptides and proteoglycolipids are inactivated or
CC deleted), a construct useful in the treatment of cancer comprising a
CC superantigen (SAg) nucleotide inserted into a virus, a mammalian T cell
CC useful in the treatment of cancer (where an adaptor protein which
CC inhibits T cell activation by tumour associated antigens is deleted or
CC functionally deactivated), a composition useful in the treatment of
CC cancer (comprising a lipid raft conjugated to a superantigen), producing
CC (M2) a tumouricidal immunocyte population ex vivo in a mammal (by
CC allowing tumour associated lipids to contact immunocytes, in which
CC receptors for the lipids are inactivated or deleted to produce a
CC tumouricidal immunocyte population, and administering the tumouricidal APC
CC activated immunocytes to the host), producing (M3) a tumour associated lipid to
CC population ex vivo in a mammal (by allowing a tumour associated lipid to
CC contact APCs, in which receptors for the tumour associated lipids are
CC inactivated or deleted to produce a tumouricidally activated population,
CC and administering APCs to the host), producing a tumouricidal T cell
CC population ex vivo in a mammal (by allowing a tumour associated lipids to
CC contact T cells, in which adaptor proteins, which inhibit T cell
CC activation by tumour associated antigens, are deleted or functionally
CC deactivated to produce a tumouricidal population of T cells, and
CC administering the tumouricidally activated T cells to the host, or
CC allowing a superantigen-lipid raft to contact T cells ex vivo, and
CC administering the tumouricidally activated T cells to the host), treating
CC (M5) cancer in a mammal (by administering a lipid binding molecule which
CC binds immunosuppressive tumour associated lipids in vivo), producing (M6)
CC a tumouricidal T cell population in vivo in a mammal (by allowing a
CC tumour associated antigen to contact immunocytes in which adaptor
CC proteins which inhibit T cell activation by tumour associated antigens
CC are deleted or functionally deactivated) and producing (M7) a
CC tumouricidal T cell population ex vivo in a mammal comprising allowing a
CC superantigen-lipid raft conjugate to contact immunocytes in vivo. The
CC receptors, methods and compositions are useful for treating cancers and
CC tumours. Bacterial superantigens are co-administered or administered as
CC fusion constructs with anti-tumour proteins or motifs. The present
CC sequence encodes an anti-tumour protein which is co-administered with or
CC incorporated into a fusion construct with a superantigen. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from the US patent
CC office website at "seqdata.uspto.gov/sequence.html?DocID=20020177551"
XX
SQ Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;
Query Match 0.4%; Score 26; DB 8; Length 17276;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
|||||
Db 6524 TTTATTGTTCAAGTGGAGAAATGAA 6549
|||||
RESULT 10
ADF43363
ID ADF43363 standard; DNA: 17276 BP.
XX
AC ADF43363;
XX
DT 12-FEB-2004 (first entry)
XX
DE Streptococcus capsular polysaccharide DNA seq id 83.
XX
XX receptor; lipid-based tumour associated antigen; cytosstatic;
XX

KM antimicrobial; gene therapy; neoplastic disease; tumour; cancer;
KM infectious disease; capsular polysaccharide; ds.
OS Streptococcus.
XX
XX
PN US2003157113-A1.
XX
XX 21-AUG-2003.
PF 28-DEC-2000; 2000US-00751708.
XX
XX 28-DEC-1999; 99US-0173371P.
XX
XX (TERM/) TERMAN D S.
XX
XX Terman DS;
XX
XX WPI; 2003-787326/74.
DR P-PSDB; ADF43364.
XX
XX
PT New receptor in a mammalian cell that inhibits regular activation by
PT receptors specific for lipid-based tumor associated antigens, useful for
PT treating a neoplastic disease or tumor, and infectious diseases.
XX
XX Example 3; SEQ ID NO 83; 151pp; English.
XX
XX The invention describes a receptor in a mammalian cell that inhibits
XX regular activation by receptors specific for lipid-based tumour
XX associated antigen. The receptor has cytostatic and antimicrobial
XX properties and is suitable for use in gene therapy. The receptors,
XX methods and compositions are useful for treating a neoplastic disease or
XX tumour (cancer), and infectious diseases. This sequence represents a
XX streptococcal capsular polysaccharide polynucleotide, a cell surface
XX moiety, the DNA of which can be transfected into a cell with superantigen
XX DNA to generate antitumour immunity.
XX
SQ Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;
Query Match 0.4%; Score 26; DB 10; Length 17276;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1755 TTTATTGTCAGTGGAGAAATGAA 1780
Db 6524 TTTATTGTCAGTGGAGAAATGAA 6549
RESULT 11
AEA03034
ID AEA03034 standard; DNA; 17276 BP.
XX
AC AEA03034;
XX
XX 28-JUL-2005 (first entry)
DT
XX
DE Staphylococcal hemolysin nucleotide sequence SEQ ID NO:60.
XX
XX tumor; neoplasm; gene therapy; immunotherapy; cytostatic; hemolysin;
KM gene; ds.
XX
XX Staphylococcus sp.
OS
XX
XX US2005112141-A1.
PN
XX
XX 26-MAY-2005.
PD
XX
PF 08-SEP-2004; 2004US-00937758.
XX
XX 30-AUG-2000; 2000US-00650884.
PR
XX (TERM/) TERMAN D S.
PA
XX
PI Terman DS;

XX
DR WPI; 2005-394926/40.
DR P-PSDB; AEA03035.
XX
XX
XX New composition for treating a tumor or neoplastic disease in a subject
PT comprises conjugates comprising superantigen polypeptides or nucleic
PT acids with other molecules that produce a tumoricidal response.
XX
XX Example 3; SEQ ID NO 60; 125pp; English.
XX
XX
XX The invention relates to a composition for treating a tumor or neoplastic
XX disease in a subject. Also described: (1) a mammalian cell comprising an
XX exogenous nucleic acid encoding a superantigen expressed in the cell,
XX which cell also produces or expresses all alpha-anomers of
XX monoglycosylceramide or diglycosylceramide, where expression of the
XX superantigen and the mono- or diglycosylceramide is capable of eliciting
XX an antitumor immune response in a mammal into which the cell is
XX introduced; (2) treating a tumor or neoplastic disease in a subject; (3)
XX preparing a population of immunotherapeutic T or natural killer T (NKT)
XX cells useful to treat a tumor or neoplastic disease in a subject; (4) an
XX apoptotic cell preparation or lysate useful for treating a tumor or
XX neoplastic disease in a subject, comprising a cell population that has
XX been transfected with naked DNA encoding a superantigen, and treated to
XX undergo apoptosis or lysis; and (5) a cell that has ingested or been
XX transfected with the above apoptotic preparation or lysate, thus,
XX rendering the cell effective in presenting material expressed from
XX a mammal to elicit an anti-tumor immune response. The composition and
XX methods are useful for treating tumors or neoplastic diseases. The
XX present sequence represents a hemolysin nucleotide sequence, which is
XX used in an example from the present invention. Note - The sequence data
XX for this patent is not represented in the printed specification, but was
XX obtained in electronic format directly from the USPTO web site.
XX
SQ Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;
Query Match 0.4%; Score 26; DB 14; Length 17276;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1755 TTTATTGTCAGTGGAGAAATGAA 1780
Db 6524 TTTATTGTCAGTGGAGAAATGAA 6549
RESULT 12
ADO40235
ID ADO40235 standard; DNA; 25020 BP.
XX
AC ADO40235;
XX
XX 15-JUL-2004 (first entry)
DT
XX
DE S. agalactiae capsular polysaccharide synthesis (cps) gene.
XX
XX Group B Streptococci; GBS; capsular polysaccharide synthesis; cps;
KM detection; ds; gene.
XX
XX Streptococcus agalactiae.
OS
XX
XX Key Location/Qualifiers
FH 153..174
FT misc_binding /*tag= a
FT /bound_moiety= "S. agalactiae cps gene specific probe"
FT 546..581
FT misc_binding /*tag= b
FT /bound_moiety= "S. agalactiae cps gene specific probe"
FT 10100..10121
FT misc_binding /*tag= c
FT /bound_moiety= "S. agalactiae cps gene specific probe"
FT 10176..10198
FT misc_binding /*tag= d
FT /bound_moiety= "S. agalactiae cps gene specific probe"

```

FT misc_binding 10219..10242
FT /tag= e
FT /bound_molecy= "S. agalactiae cps gene specific probe"
XX
XX
XX US2004009574-A1.
XX
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2002; 2002US-00192280.
XX
XX 09-JUL-2002; 2002US-00192280.
XX
XX (DATT/) DATTAGUPTA N.
XX (SHAH/) SHAH K.
XX
XX Datagupta N, Shah K;
XX
XX WPI; 2004-090471/09.
XX DR GENBANK; AFB028896.
XX
XX Novel oligonucleotide probes for detecting group B Streptococci e.g.
XX Streptococcus agalactiae in samples.
XX
XX Example 1; Fig 1; 28pp; English.
XX
XX The invention relates to compositions and methods for detecting group B
XX Streptococci (GBS) Streptococcus agalactiae capsular polysaccharide
XX synthesis (cps) gene in sample. The invention is useful for detecting S.
XX agalactiae in a sample. The present sequence is S. agalactiae cps gene.
XX This sequence is used to illustrate the method of the invention.
XX
XX Sequence 25020 BP; 8592 A; 3630 C; 4594 G; 8204 T; 0 U; 0 Other;
XX
XX Query Match 0.4%; Score 26; DB 12; Length 25020;
XX Best Local Similarity 100.0%; Pred. No. 1.8;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1755 TTTATTGTTCACTGGAGAAATGAA 1780
XX 6914 TTTATTGTTCACTGGAGAAATGAA 6939
XX
XX RESULT 13
XX ADV87741
XX ID ADV87741 standard; DNA; 95596 BP.
XX
XX ADV87741;
XX
XX 24-FEB-2005 (first entry)
XX
XX Streptococcus agalactiae DNA sequence, SEQ ID 135.
XX
XX Antibacterial; Vaccine; bacterial infection; ds.
XX
XX Streptococcus agalactiae.
XX
XX FR824074-A1.
XX
XX 31-OCT-2002.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus

```

```

PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
XX Claim 1; SEQ ID NO 135; 2687pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
XX ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
XX agalactiae involved in the synthesis of amino acids, cell membranes,
XX intermediate (central) metabolism, energetic metabolism, fatty acid and
XX phospholipid metabolism, nucleotide metabolism including purines,
XX pyrimidines and/or nucleosides, regulatory functions, replication,
XX transcription, translation, protein transport, adaptation to atypical
XX conditions, sensitivity to medicines and/or analogues, functions related
XX to transposons, biosynthesis of cofactors, prosthetic groups and
XX to transporters, cell membrane proteins and cellular machinery. (I) are
XX useful for the detection and/or amplification of nucleic acids.
XX CC Pharmaceutical composition comprising (I) or (II) are useful for
XX treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is
XX equivalent for the present basic patent FR2824074A1. WO200292818A2
XX contains 6617 sequence whereas the present patent only contains 2344
XX sequences.
XX
XX Sequence 95596 BP; 31762 A; 16284 C; 17028 G; 30522 T; 0 U; 0 Other;
XX
XX Query Match 0.4%; Score 26; DB 13; Length 95596;
XX Best Local Similarity 100.0%; Pred. No. 1.6;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1755 TTTATTGTTCACTGGAGAAATGAA 1780
XX 5332 TTTATTGTTCACTGGAGAAATGAA 5357
XX
XX RESULT 14
XX ADV8994
XX ID ADV8994 standard; DNA; 95596 BP.
XX
XX ADV8994;
XX
XX 24-FEB-2005 (first entry)
XX
XX Streptococcus agalactiae DNA sequence, SEQ ID 135.
XX
XX Antibacterial; vaccine; bacterial infection; ds.
XX
XX Streptococcus agalactiae.
XX
XX WO200292818-A2.
XX
XX 21-NOV-2002.
XX
XX 26-APR-2002; 2002WO-IB003059.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
XX
XX Claim 1; SEQ ID NO 135; 439pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV8860-ADV8998 and ADV8341-ADV85476) and
XX novel polypeptides (II; ADV8999-ADV81203 and ADV81205-ADV83340). The

```


CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons, biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (i) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (i) or (ii) are
CC useful for treatment of a bacterial *S. agalactiae* infection. The complete
CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
CC contains only 2344 sequences.

XX
SQ Sequence 95596 BP; 31762 A; 16282 C; 17030 G; 30522 T; 0 U; 0 Other;

Query Match 0.4%; Score 26; DB 13; Length 95596;

Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTCAGTGGGAAGAATGAA 1780
|||||
DB 5332 TTTATTGTCAGTGGGAAGAATGAA 5357

RESULT 15

ABN71527_11/c

Continuation (12 of 22) of ABN71527 from base 1100001 (*Streptococcus polynucleotide* SEQ
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527

WP	Fragment Name	Begin	End
WP	ABN71527_00	1	110000
WP	ABN71527_01	100001	210000
WP	ABN71527_02	200001	310000
WP	ABN71527_03	300001	410000
WP	ABN71527_04	400001	510000
WP	ABN71527_05	500001	610000
WP	ABN71527_06	600001	710000
WP	ABN71527_07	700001	810000
WP	ABN71527_08	800001	910000
WP	ABN71527_09	900001	1010000
WP	ABN71527_10	1000001	1110000
WP	ABN71527_11	1100001	1210000
WP	ABN71527_12	1200001	1310000
WP	ABN71527_13	1300001	1410000
WP	ABN71527_14	1400001	1510000
WP	ABN71527_15	1500001	1610000
WP	ABN71527_16	1600001	1710000
WP	ABN71527_17	1700001	1810000
WP	ABN71527_18	1800001	1910000
WP	ABN71527_19	1900001	2010000
WP	ABN71527_20	2000001	2110000
WP	ABN71527_21	2100001	2155561

Query Match 0.4%; Score 26; DB 6; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTCAGTGGGAAGAATGAA 1780
|||||
DB 44803 TTTATTGTCAGTGGGAAGAATGAA 44778

Search completed: December 25, 2005, 18:09:20
Job time : 2510 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 16:25:16 / Search time 16548 Seconds
(without alignments)
19768.871 Million cell updates/sec

Title: US-09-767-041-9

Perfect score: 6992
Sequence: 1 atcgccaaagaattgca.....gcaaacagcaagaagctt 6992

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched: 41078325 seqs, 23393541228 residues

Word size: 0

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss81:*
10: gb_gss82:*
11: gb_gss83:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	0.4	577	10	CM174539
2	25	0.4	408	8	BB786691
3	25	0.4	408	8	DR923160
4	25	0.4	408	8	DR923160
5	24	0.3	175	1	AV341783
6	24	0.3	222	1	AV365377
7	24	0.3	228	1	AA571514
8	24	0.3	242	1	AV243993
9	24	0.3	251	1	AV278884
10	24	0.3	258	1	AV314331
11	24	0.3	269	1	AV060017
12	24	0.3	277	1	AV296670
13	24	0.3	292	1	AV225401
14	24	0.3	299	1	AV014839
15	24	0.3	304	1	AV030767
16	24	0.3	328	1	AV351151
17	24	0.3	359	1	AV397025
18	24	0.3	361	1	AI847613
19	24	0.3	368	5	BY674466
20	24	0.3	371	5	BY404552
21	24	0.3	382	1	AI121971
22	24	0.3	384	3	BM210733

C 23	24	0.3	395	6	CD564418	CD564418	B0478E09-
C 24	24	0.3	397	5	BY582580	BY582580	BY582580
C 25	24	0.3	397	5	BY679597	BY679597	BY679597
C 26	24	0.3	405	5	BY671706	BY671706	BY671706
C 27	24	0.3	408	2	BB789287	BB789287	BB789287
C 28	24	0.3	411	1	AI840087	AI840087	AI840087
C 29	24	0.3	426	1	AA111987	AA111987	AA111987
C 30	24	0.3	426	2	BB784004	BB784004	BB784004
C 31	24	0.3	431	1	AA084098	AA084098	AA084098
C 32	24	0.3	439	1	BY484542	BY484542	BY484542
C 33	24	0.3	445	2	BB823210	BB823210	BB823210
C 34	24	0.3	455	1	AA067354	AA067354	AA067354
C 35	24	0.3	466	1	AI606961	AI606961	AI606961
C 36	24	0.3	490	2	BE951202	BE951202	BE951202
C 37	24	0.3	490	6	CF540840	CF540840	CF540840
C 38	24	0.3	533	5	BY412569	BY412569	BY412569
C 39	24	0.3	587	3	BM241066	BM241066	BM241066
C 40	24	0.3	600	5	BU923617	BU923617	BU923617
C 41	24	0.3	618	6	CF469373	CF469373	CF469373
C 42	24	0.3	622	3	BI653525	BI653525	BI653525
C 43	24	0.3	640	7	CK621720	CK621720	CK621720
C 44	24	0.3	669	7	CV517509	CV517509	CV517509
C 45	24	0.3	669	7	CV517509	CV517509	CV517509

ALIGNMENTS

RESULT 1
LOCUS CM174539
DEFINITION 104 586 11157462 148 36560 019 SORGHUM METHYLATION FILTERED LIBRARY (LIDID: 104) SORGHUM bicolor genomic clone 11157462, genomic survey

SEQUENCE:
CM174539 577 bp DNA linear GSS 29-OCT-2004
CM174539 104 586 11157462 148 36560 019 SORGHUM METHYLATION FILTERED LIBRARY (LIDID: 104) SORGHUM bicolor genomic clone 11157462, genomic survey
CM174539 GI:54867106
VERSION
KEYWORDS
SOURCE
ORGANISM
SORGHUM bicolor (sorghum)
SORGHUM bicolor
Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE
AUTHORS
Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holeman, H., Roe, B.A., Wiley, G., Korf, I.F., Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and Martienssen, R.A.

REFERENCE
AUTHORS
Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holeman, H., Roe, B.A., Wiley, G., Korf, I.F., Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and Martienssen, R.A.

REFERENCE
AUTHORS
Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holeman, H., Roe, B.A., Wiley, G., Korf, I.F., Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and Martienssen, R.A.

TITLE
JOURNAL
PUBMED
COMMENT
15660154
Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: 586 Row: m Column: 06
Seq primer: SWfor Forward
Class: methylation filtered
High quality sequence stop: 577.
Location/Qualifiers
1. 577
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/cultiivar="ATx623"
/db_xref="taxon:4558"
/clone="11157462"
/clone_lib="Sorghum methylation filtered library (LIDID: 104)"
/note="Organ: leaf, Vector: pBCSK(-); Site: 1; HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5

ORIGIN

kb fraction, ligated into HincII-digested pBSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

Query Match 0.4%; Score 26; DB 10; Length 577;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2801 TTATGAAAAATGATTTTCAAT 2826
|||||
DB 235 TTATGAAAAATGATTTTCAAT 260

RESULT 2

BB786691/c 408 bp mRNA linear EST 08-JUL-2003
BB786691 RIKEN full-length enriched, kidney CCL-142 RAG cDNA Mus
LOCUS BB786691
DEFINITION musculus cDNA clone G430113A03 3', mRNA sequence.
ACCESSION BB786691
VERSION BB786691.1 GI:16955187
KEYWORDS EST.

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Mus.
1 (bases 1 to 408)

REFERENCE

AUTHORS

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akanira,S., Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A., Muramatsu,M., and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL

COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsuhira,S., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y., and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES

SOURCE

Location/Qualifiers
1..408
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"

ORIGIN

Query Match 0.4%; Score 25; DB 2; Length 408;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3605 ATTTTCTTCTGTCCTGCTTTT 3629
|||||
DB 384 ATTTTCTTCTGTCCTGCTTTT 360

RESULT 3

DR923160/c 419 bp mRNA linear EST 02-AUG-2005
DR923160 Aquilegia cDNA library Aquilegia formosa x Aquilegia
LOCUS DR923160
DEFINITION pubescens cDNA clone CO1MU36, mRNA sequence.
ACCESSION DR923160
VERSION DR923160.1 GI:71692523
KEYWORDS EST.

SOURCE Aquilegia formosa x Aquilegia pubescens
ORGANISM Aquilegia formosa x Aquilegia pubescens

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales; Ranunculaceae; Aquilegia.
1 (bases 1 to 419)

REFERENCE

Hodges,S.A., Rensink,M., Buell,C.R., Borevitz,J., Kramer,E., Nordborg,M., and Tomkins,J.

JOURNAL

Unpublished (2005)
Other ESTs: EST1114700

COMMENT

Contact: Scott Hodges
Department of Ecology, Evolution and Marine Biology
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
Tel: 805 893 7813
Fax: 805 893 4724
Email: hodges@lifesci.ucsb.edu
Seq primer: TTTTCTTCTTCTTCTTCTTCTT (where N = A, G & C).

FEATURES

SOURCE

Location/Qualifiers
1..419
/organism="Aquilegia formosa x Aquilegia pubescens"
/mol_type="mRNA"
/db_xref="taxon:338618"
/clone="CO1MU36"
/tissue_type="mixed shoot and floral apical meristems, flower buds, leaves and roots"
/lab_host="B103 T1 (T1 and T5 phage resistance)"
/clone_lib="Aquilegia cDNA library"
/note="Vector: pCMV SPORT.1; Site: EcoRI, Site 2: NotI; F2, F3, and F4 lines of Aquilegia formosa x A. pubescens were grown from seed in greenhouses at UC Santa Barbara. From these plants three sets of tissue were collected: 1) Small flower buds (<10 mm) and very young inflorescences (T1 & T2) by weight respectively, 2) Medium (7-20 mm) and large (at or near anthesis) flower buds (65 & 35% by weight respectively) and 3) Shoot apical meristems. A fourth set of tissue was collected from plants of A. formosa. These plants were grown from seed in sand and at approximately 1 month root tissue and leaf tissue of various developmental stages were collected (64 & 16% by weight respectively). Total RNA was extracted from each set of tissue and pooled in the following proportions: 1) 1.5x from sets 1 & 2, 1x from sets 3 & 4. From the pooled total RNA, mRNA was extracted and enriched for full-length mRNAs and then normalized with proprietary methods by Invitrogen."

ORIGIN	Query Match	0.4%	Score 25;	DB 8;	Length 419;
	Best Local Similarity	100.0%	Pred. No. 13;		
	Matches	25;	Conservative 0;	Mismatches 0;	Indels 0;
OY	5447	TGTTGTGATAGTAGATTAGTAGAA	5471		
DB	134	TGTTGTGATAGTAGATTAGTAGAA	110		
RESULT 4					
BH304568					
LOCUS	BH304568	517 bp	DNA	linear	GSS 30-NOV-2001
DEFINITION	CH230-179M24.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone				
ACCESSION	CH230-179M24				
VERSION	BH304568				
KEYWORDS	BH304568.1 GI:17216976				
SOURCE	GSS.				
ORGANISM	Rattus norvegicus (Norway rat)				
REFERENCE	Rattus norvegicus				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Rattus.				
TITLE	1 (bases 1 to 517)				
JOURNAL	Zhao,S., Shetty,J., Shatsman,S., Tsengaye,G., Geer,K., Shwartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,P., de Jong,P. and Fraser,C.M.				
COMMENT	Rat BAC End Sequences from Library CHORI-230 EcoRI segment Unpublished (1999)				
	Other_GSSs: CH230-179M24.TV				
	Contact: Shaying Zhao				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: szhao@tigr.org				
	Clones are derived from the rat BAC library CHORI-230				
	(http://www.chori.org/bacpac/cb230.htm). For BAC library				
	availability, please contact Plecter de Jong (pdejong@tigr.org).				
	Clones may be purchased from BACPAC Resources				
	(http://www.chori.org/bacpac/or_oring_information.htm). BAC end				
	page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html				
	Plate: 179 row: M column: 24				
	Seq primer: T7				
	Class: BAC ends.				
FEATURES	Location/Qualifiers				
Source	1..517				
	/organism="Rattus norvegicus"				
	/mol_type="genomic DNA"				
	/strain="BN/SaNHsd/MCM"				
	/db_xref="taxon:10116"				
	/clone="CH230-179M24"				
	/sex="Female"				
	/cell_type="Brain"				
	/clone_1ib="CHORI-230 Segment 1"				
	/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;				
	CHORI-230 Rat (BN/SaNHsd/MCM) BAC library produced by				
	Plecter de Jong"				
ORIGIN					
Query Match	0.4%	Score 25;	DB 9;	Length 517;	
Best Local Similarity	100.0%	Pred. No. 13;			
Matches	25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	2154	TTTTCAGATGATATATATTTTATT	2178		
DB	396	TTTTCAGATGATATATATTTTATT	420		

LOCUS	AVJ41783	175 bp	mRNA	linear	EST 11-NOV-1999
DEFINITION	AVJ41783 RIKEN full-length enriched, adult male olfactory bulb musculus cDNA clone 6430524F20 3' similar to AF067395 Mus musculus Nix (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.				
ACCESSION	AVJ41783				
VERSION	AVJ41783.1	GI:6381835			
KEYWORDS	EST,				
ORGANISM	Mus musculus (house mouse)				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus. 1 (bases 1 to 175)				
TITLE	Komno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukuishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Iribikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,T., Kikuchi,N., Kojima,Y., Koya,S., Kuwabake,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sanio,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogaue,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateono,M., Tomimaga,N., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.				
JOURNAL COMMENT	RIKEN Mouse ESTs (Komno,H., et al. 1999) Unpublished (1999) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.jp, URL:http://genome.gsc.riken.jp/ Sasaki,N., Izawa,M., Watabiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998) Itou,M., Kitsuai,T., Akiyama,Y., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.				
FEATURES	Location/Qualifiers				
source	1..175 organism="Mus musculus" mol_type="mRNA" strain="CS7BL/6J" db_xref="taxon:10090" clone="6430524F20" sex="male" tissue_type="olfactory brain" dev_stage="adult" lab_host="DH10B" clone_id="RIKEN full-length enriched, adult male olfactory bulb" note="Site 1: SalI, Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed by a primer [5'-GAGGAGAAGAGATCCAGAGCTCTTTTTCCTTTTTTTTNN 3'-], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization				

source 1. .228

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J x DBA/2J F1"

/db_xref="taxon:10090"

/clone_xref="IMAGE:989348"

/tissue_type="blastocyst"

/dev_stage="embryo (pre-implantation)"

/lab_host="DH10B"

/clone_11b="Knowles Solter mouse blastocyst B1"

/note="Organ: embryo, Vector: pSPORT, Site_1: NotI; Site_2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(dT): 5'-CGGCGACCGCGACGCTTTTCTTTT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies). Two different size selections: B1 (larger inserts) and B3."

ORIGIN

Query Match 0.3%; Score 24; DB 1; Length 228;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3606 TTTTCTTCTGTCCTGCTTT 3629

101 TTTTCTTCTGTCCTGCTTT 78

RESULT 8

AV243993 242 bp mRNA linear EST 04-NOV-1999

LOCUS AV243993 RIKEN full-length enriched, 0 day neonate head Mus

DEFINITION musculus cDNA clone 4831426M13 3' similar to AF067395 Mus musculus NIX (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.

ACCESSION AV243993

VERSION AV243993.1 GI:6231452

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 242)

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomlinaga, N., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. RIKEN Mouse ESTs (Kono, H., et al. 1999)

Unpublished (1999)

Contact: Yoshinori Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/

Sasaki, N., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)

Itoh, M., Katsunari, P., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

source 1. .242

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_xref="IMAGE:989348"

/tissue_type="blastocyst"

/dev_stage="embryo (pre-implantation)"

/lab_host="DH10B"

/clone_11b="RIKEN full-length enriched, 0 day neonate head"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCAGCTATTAATTAATTAATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from lambda FLC I."

ORIGIN

Query Match 0.3%; Score 24; DB 1; Length 242;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3606 TTTTCTTCTGTCCTGCTTT 3629

214 TTTTCTTCTGTCCTGCTTT 191

RESULT 9

AV278884 251 bp mRNA linear EST 05-NOV-1999

LOCUS AV278884 RIKEN full-length enriched, adult male testis (DH10B) Mus

DEFINITION musculus cDNA clone 4933404118 3' similar to AF067395 Mus musculus NIX (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.

ACCESSION AV278884

VERSION AV278884.1 GI:6266921

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 251)

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomlinaga, N., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. RIKEN Mouse ESTs (Kono, H., et al. 1999)

Unpublished (1999)

Contact: Yoshinori Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/

Sasaki, N., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)

Itoh, M., Katsunari, P., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

TITLE
JOURNAL
COMMENT

RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/
Sasaki, N., Izawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsunaga, S., Carninci, P., Muramatsu, M., Okazaki, Y., and
Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Kitanai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., and Hayashizaki, Y.
Okazaki, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.irc.riken.go.jp) for
further details.

FEATURES
source

Location/Qualifiers
1..251
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4933404118"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_11b="RIKEN full-length enriched, adult male testis
(DH10B)"

/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATCTCCAGATTATTAATTAATCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified Bluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."

ORIGIN

Query Match 0.3%; Score 24; DB 1; Length 251;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3606 TTTTCTCTGTCCTGCTTT 3629
Db 218 TTTTCTCTGTCCTGCTTT 195

RESULT 10
AV314331/c 258 bp mRNA linear EST 08-NOV-1999
LOCUS AV314331 RIKEN full-length enriched, adult male thymus Mus musculus
DEFINITION cDNA clone 5830417011 3' similar to AF067395 Mus musculus NIX (NIX)
mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.
ACCESSION AV314331
VERSION AV314331.1 GI:6279583

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 258)
Konno, H., Aizawa, K., Akhita, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I.,
Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsumura, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,
Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomimaga, N.,
Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A.,
Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/
Sasaki, N., Izawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsunaga, S., Carninci, P., Muramatsu, M., Okazaki, Y., and
Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Kitanai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
Okazaki, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.irc.riken.go.jp) for
further details.

FEATURES
source

Location/Qualifiers
1..258
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5830417011"
/sex="male"
/tissue_type="thymus"
/dev_stage="adult"
/lab_host="DH10B"
/clone_11b="RIKEN full-length enriched, adult male thymus"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA was through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATCTCCAGATTATTAATTAATCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified Bluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
BamHI."

ORIGIN

RESULT 12
AV296670/c

primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATTAATCCCCCCCCCCCC
3']. cDNA

GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a

ORIGIN

modified pBluescript KS(+) after bulk excision from Lambda
 FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

Query Match 0.3%; Score 24; DB 1; Length 277;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3606 TTTTTCCTGTCCTGCTTTT 3629
 ||||||||||||||||||||
 Db 242 TTTTTCCTGTCCTGCTTTT 219

RESULT 13
 AV225401/c
 LOCUS AV225401 292 bp mRNA linear EST 14-NOV-2001
 DEFINITION AV225401 RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue Mus musculus cDNA clone 3830429H23 3' similar to AF067395 Mus musculus NIX (Nix) mRNA, nuclear gene encoding mitochondrial protein, mRNA sequence.

ACCESSION AV225401
 VERSION AV225401.1 GI:6176716

KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS 1 (bases 1 to 292)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koye, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomimaga, N., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamuro, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al. 1999)

TITLE Unpublished (1999)
 JOURNAL Contact: Yoshihide Hayashizaki
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9226
 Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/
 Sasaki, N., Izawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

source Location/Qualifiers
 1..292
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="3830429H23"
 /sex="female"
 /tissue_type="placenta and extra embryonic tissue"

ORIGIN

Query Match 0.3%; Score 24; DB 1; Length 292;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3606 TTTTTCCTGTCCTGCTTTT 3629
 ||||||||||||||||||||
 Db 260 TTTTTCCTGTCCTGCTTTT 237

RESULT 14
 AV014839/c
 LOCUS AV014839 299 bp mRNA linear EST 25-AUG-1999
 DEFINITION AV014839 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA clone 1110056J23, mRNA sequence.

ACCESSION AV014839
 VERSION AV014839.1 GI:4791831
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE

AUTHORS 1 (bases 1 to 299)
 Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Horii, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Natsuna, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamuro, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 RIKEN Mouse ESTs

TITLE Unpublished (1999)
 JOURNAL RIKEN Mouse ESTs
 COMMENT Contact: Chie Owa

Genome Science Laboratory
 RIKEN
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@sc.riken.go.jp

Thermolabile and thermostable activation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES

source Location/Qualifiers
 1..299
 /organism="Mus musculus"

```

/mol_type="mRNA"
/strain="C57BL/6J"
/ld_xref="taxon:10090"
/clone="1110056J23"
/sex="mixed"
/dev_stage="18-day embryo"
/clone_lib="Mus musculus 18-day embryo C57BL/6J"

```

Query	March	0.3%	Score 24;	DB 1;	Length 299;
Best Local Similarity		100.0%;	Pred. No. 39;		
Matches	24;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	3606	TTTTTTTCTGTCGTCGCTGCTTT	3629		
Db	273	TTTTTTTCTGTCGTCGCTGCTTT	250		

RESULT 15	AV303767/c	LOCUS	304 bp	mRNA	linear	EST 10-NOV-1999
DEFINITION	AV303767	RIKEN full-length enriched, 8 days embryo Mus musculus				

ACCESSION	AV303767
VERSION	AV303767.1
	GI:6336281

SOURCE ORGANISM	Mus musculus (house mouse)	Mus musculus
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

REFERENCE
AUTHORS

TITLE	RIKEN Mouse ESTs (Kono, H., et al. 1999)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Yoshihide Hayashizaki

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan
Tel.: 81-45-503-9232
Fax: 81-45-503-9216
Email: genome-teseq@ic.riken.jp, URL: <http://genome.gsc.riken.jp/>
Sasaki, N., Iizawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and
Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)
Itoh, M., Kikunishi, T., Akiyama, Y., Shibata, K., Iizawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
Okazaki, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
13-44 (1999)
Please visit our web site (<http://genome.irc.riken.go.jp>) for
further details.

```
FEATURES
source      location/qualifiers
1..304      /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
```

QY	3606	TTTTTTCTGTGTGCTTGCCTTT	3629
DB	270	TTTTTTCTGTGTGCTTGCCTTT	247

Query Match	Best Local Similarity	100.0%	Pred. No. 39,
Matches	24;	Conservative	0;
Mismatches			0;
Indels			0;
Gaps			0;
Score	24;	DB 1;	Length 304;

ORIGIN
<p> //db_xref="taxon:10090" //clone="5730515C07" //sex="mixed" //dev_stage="8 days embryo" //lab_host="DR108" //clone_1lb="RIKEN full-length enriched, 8 days embryo" //note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluescript KS(+) after bulk excision from LambdaFLC 1. Cloning sites, 5' end: SalI, 3' end: BamHI." </p>

Search completed: December 26, 2005, 05:00:53
Job time : 16553 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 17:21:11 ; Search time 781 Seconds

(without alignments)
15913.843 Million cell updates/sec

Title: US-09-767-041-9

Perfect score: 6992
Sequence: 1 atcgcaacgaattgca.....gcaacagcaaaagactt 6992

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 1303057 seqs, 888780828 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/1 COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5 COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/H COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/PC/US COMB.seq:*
- 7: /cgn2_6/prodata/1/ina/PP COMB.seq:*
- 8: /cgn2_6/prodata/1/ina/RE COMB.seq:*
- 9: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	0.4	6865	US-09-900-038A-3	Sequence 3, Appl1
2	24	0.3	142783	US-09-949-016-15127	Sequence 15127, A
3	23	0.3	547	US-09-270-767-11261	Sequence 11261, A
4	23	0.3	221545	US-09-949-016-13875	Sequence 13875, A
5	23	0.3	580073	US-08-545-528D-1	Sequence 1, Appl1
6	22	0.3	579	US-09-328-352-3168	Sequence 1368, Ap
7	22	0.3	66164	US-09-949-016-13574	Sequence 13574, A
8	21	0.3	180	US-09-270-767-26807	Sequence 26807, A
9	21	0.3	601	US-09-949-016-153210	Sequence 153210,
10	21	0.3	601	US-09-949-016-195516	Sequence 195516,
11	21	0.3	601	US-09-949-016-195517	Sequence 195517,
12	21	0.3	1143	US-09-543-681A-1740	Sequence 1740, Ap
13	21	0.3	1170	US-09-495-406-11	Sequence 11, Appl
14	21	0.3	1170	US-09-816-028A-15	Sequence 15, Appl
15	21	0.3	1170	US-10-303-162-15	Sequence 15, Appl
16	21	0.3	1170	US-10-303-134-15	Sequence 15, Appl
17	21	0.3	1170	US-10-303-118-15	Sequence 15, Appl
18	21	0.3	1170	US-10-303-128-15	Sequence 15, Appl
19	21	0.3	1845	US-08-956-171E-525	Sequence 525, App
20	21	0.3	1845	US-08-781-986A-525	Sequence 525, App
21	21	0.3	2145	US-09-854-300-5	Sequence 5, Appl1
22	21	0.3	4557	US-08-956-171E-341	Sequence 341, App
23	21	0.3	4557	US-08-781-986A-341	Sequence 341, Appl
24	21	0.3	11474	US-09-495-406-1	Sequence 1, Appl1

25	21	0.3	11474	US-09-816-028A-1	Sequence 1, Appl1
26	21	0.3	11474	US-10-303-162-1	Sequence 1, Appl1
27	21	0.3	11474	US-10-303-134-1	Sequence 1, Appl1
28	21	0.3	11474	US-10-303-118-1	Sequence 1, Appl1
29	21	0.3	11474	US-10-303-128-1	Sequence 1, Appl1
30	21	0.3	16442	US-08-781-891-208	Sequence 208, App
31	21	0.3	16442	US-09-618-166-208	Sequence 208, App
32	21	0.3	32392	US-09-662-254B-27	Sequence 27, Appl
33	21	0.3	87644	US-09-949-016-16041	Sequence 16041, A
34	21	0.3	97221	US-09-949-016-12755	Sequence 12755, A
35	21	0.3	112114	US-09-949-016-17592	Sequence 17292, A
36	21	0.3	129554	US-09-949-002-765	Sequence 765, App
37	21	0.3	154605	US-09-949-016-11894	Sequence 11894, A
38	21	0.3	194915	US-09-949-016-15584	Sequence 15584, A
39	21	0.3	374159	US-09-949-016-15868	Sequence 15868, A
40	21	0.3	393753	US-09-949-016-14573	Sequence 14573, A
41	21	0.3	393753	US-09-949-016-14574	Sequence 14574, A
42	21	0.3	421118	US-09-949-016-16297	Sequence 16297, A
43	21	0.3	640681	US-09-790-988-1	Sequence 1, Appl1
44	21	0.3	818128	US-09-949-016-14546	Sequence 14546, A
45	21	0.3	818128	US-09-949-016-14547	Sequence 14547, A

ALIGNMENTS

```
RESULT 1
US-09-900-038A-3
; Sequence 3, Application US/09900038A
; Patent No. 6828128
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsuhide
; APPLICANT: Matanabe, Masaki
; APPLICANT: Iijima, Shuji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 766,53
; CURRENT APPLICATION NUMBER: US/09/900,038A
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: JP 2001-392
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6865
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae type Ib
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (617)..(1789)
; NAME/KEY: CDS
; LOCATION: (1816)..(2262)
; NAME/KEY: CDS
; LOCATION: (2265)..(2744)
; NAME/KEY: CDS
; LOCATION: (2843)..(3979)
; NAME/KEY: CDS
; LOCATION: (3982)..(4953)
; NAME/KEY: CDS
; LOCATION: (5009)..(5947)
; NAME/KEY: CDS
; LOCATION: (5009)..(5947)
; US-09-900-038A-3

Query Match          0.4%; Score 26; DB 3; Length 6865;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1755 TTTATTGTTTCAGTGGAGAAATGAA 1780
|||||
Db 2197 TTTATTGTTTCAGTGGAGAAATGAA 2222
|||||

```
RESULT 2
US-09-949-016-15127
; Sequence 15127, Application US/09949016
```

```
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15127
/ LENGTH: 142783
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(142783)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15127

Query Match          0.3%; Score 24; DB 3; Length 142783;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2959 TTTAAATAGAAATTAATCTAAA 2982
DB      141086 TTTAAATAGAAATTAATCTAAA 141109

RESULT 3
US-09-270-767-11261/C
/ Sequence 11261, Application US/09270767
/ Patent No. 6703491
/ GENERAL INFORMATION:
/ APPLICANT: Homburger et al.
/ TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
/ FILE REFERENCE: File Reference: 7326-094
/ CURRENT APPLICATION NUMBER: US/09/270,767
/ CURRENT FILING DATE: 1999-03-17
/ NUMBER OF SEQ ID NOS: 62517
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 11261
/ LENGTH: 547
/ TYPE: DNA
/ ORGANISM: Drosophila melanogaster
US-09-270-767-11261

Query Match          0.3%; Score 23; DB 3; Length 547;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5630 TTTGTATATAATATATTTT 5652
DB      105 TTTGTATATAATATATTTT 83

RESULT 4
US-09-949-016-13875
/ Sequence 13875, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
```

```
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13875
/ LENGTH: 221545
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-13875

Query Match          0.3%; Score 23; DB 3; Length 221545;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3189 AAAATGAAATTTATGTTT 3211
DB      203243 AAAATGAAATTTATGTTT 203265

RESULT 5
US-08-545-528D-1
/ Sequence 1, Application US/08545528D
/ Patent No. 6537773
/ GENERAL INFORMATION:
/ APPLICANT: Fraser et al.
/ TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
/ Patent No. 6537773
/ TITLE OF INVENTION: Thereof, and Uses Thereof
/ FILE REFERENCE: PB193P1
/ CURRENT APPLICATION NUMBER: US/08/545,528D
/ CURRENT FILING DATE: 1995-10-19
/ PRIOR APPLICATION NUMBER: US 08/488,018
/ PRIOR FILING DATE: 1995-06-07
/ PRIOR APPLICATION NUMBER: US 08/473,545
/ PRIOR FILING DATE: 1995-06-07
/ NUMBER OF SEQ ID NOS: 1
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 580073
/ TYPE: DNA
/ ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match          0.3%; Score 23; DB 3; Length 580073;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2981 AATTTAAGAAATTATCTAAA 3003
DB      534125 AATTTAAGAAATTATCTAAA 534147

RESULT 6
US-09-328-352-3168
/ Sequence 3168, Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 3168
/ LENGTH: 579
/ TYPE: DNA
/ ORGANISM: Acinetobacter baumannii
US-09-328-352-3168
```


Query Match 0.3%; Score 22; DB 3; Length 579;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5244 ATTAATCAAAATATATAATA 5265
|||||
DB 85 ATTAATCAAAATATATAATA 106

RESULT 7
US-09-949-016-13574
; Sequence 13574, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13574
; LENGTH: 66164
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13574

Query Match 0.3%; Score 22; DB 3; Length 66164;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1997 GAACATATATTAACAATCAG 2018
|||||
DB 63740 GAACATATATTAACAATCAG 63761

RESULT 8
US-09-270-767-26807/c
; Sequence 26807, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 26807
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-26807

Query Match 0.3%; Score 21; DB 3; Length 180;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5630 TTGTATATAATATATTTT 5650
|||||
DB 21 TTGTATATAATATATTTT 1

RESULT 9
US-09-949-016-153210

; Sequence 153210, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153210
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-153210

Query Match 0.3%; Score 21; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4334 AGAAATTCCTTAACATATA 4354
|||||
DB 481 AGAAATTCCTTAACATATA 501

RESULT 10
US-09-949-016-195516/c
; Sequence 195516, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195516
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195516

Query Match 0.3%; Score 21; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3138 ATGATTAATTTATTTTAT 3158
|||||
DB 375 ATGATTAATTTATTTTAT 355

RESULT 11
US-09-949-016-195517/c
; Sequence 195517, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 195517
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
/ US-09-949-016-195517

Query Match          0.3%; Score 21; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3138 ATGAAATTTAATTTATTTCT 3158
DB 582 ATGAAATTTAATTTATTTCT 562

RESULT 12
US-09-543-681A-1740
/ Sequence 1740, Application US/09543681A
/ Patent No. 6605709
/ GENERAL INFORMATION:
/ APPLICANT: GARY BRETON
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
/ FILE REFERENCE: 2709,1002-001
/ CURRENT APPLICATION NUMBER: US/09/543,681A
/ CURRENT FILING DATE: 2000-04-05
/ PRIOR APPLICATION NUMBER: US 60/128,706
/ PRIOR FILING DATE: 1999-04-09
/ NUMBER OF SEQ ID NOS: 8344
/ SEQ ID NO 1740
/ LENGTH: 1143
/ TYPE: DNA
/ ORGANISM: Proteus mirabilis
/ US-09-543-681A-1740

Query Match          0.3%; Score 21; DB 3; Length 1143;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2975 AATCTAAATTTAAGAAATTA 2995
DB 137 AATCTAAATTTAAGAAATTA 157

RESULT 13
US-09-495-406-11
/ Sequence 11, Application US/09495406
/ Patent No. 6503744
/ GENERAL INFORMATION:
/ APPLICANT: Gilbert, Michel
/ APPLICANT: Wakarchuk, Warren W.
/ APPLICANT: National Research Council of Canada
/ TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
/ FILE REFERENCE: 019633-000110US
/ CURRENT APPLICATION NUMBER: US/09/495,406
/ CURRENT FILING DATE: 2000-01-31
/ PRIOR APPLICATION NUMBER: US 60/118,213
/ PRIOR FILING DATE: 1999-02-01
/ NUMBER OF SEQ ID NOS: 35
/ SOFTWARE: PatentIn Ver. 2.1
```

```
/ SEQ ID NO 11
/ LENGTH: 1170
/ TYPE: DNA
/ ORGANISM: Campylobacter jejuni
/ FEATURE:
/ OTHER INFORMATION: glycosyltransferase from C. jejuni strain OH4384
/ US-09-495-406-11

Query Match          0.3%; Score 21; DB 3; Length 1170;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6195 TATTAATGTGAAAAATATTTA 6215
DB 31 TATTAATGTGAAAAATATTTA 51

RESULT 14
US-09-816-028A-15
/ Sequence 15, Application US/09816028A
/ Patent No. 6699705
/ GENERAL INFORMATION:
/ APPLICANT: Gilbert, Michel
/ APPLICANT: Wakarchuk, Warren W.
/ APPLICANT: National Research Council of Canada
/ TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
/ FILE REFERENCE: 019633-00011US
/ CURRENT APPLICATION NUMBER: US/09/816,028A
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: US 60/118,213
/ PRIOR FILING DATE: 1999-02-01
/ PRIOR APPLICATION NUMBER: US 09/495,406
/ PRIOR FILING DATE: 2000-01-31
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 15
/ LENGTH: 1170
/ TYPE: DNA
/ ORGANISM: Campylobacter jejuni
/ FEATURE:
/ OTHER INFORMATION: glycosyltransferase from C. jejuni strain OH4384
/ OTHER INFORMATION: (ORF 4a of lipooligosaccharide (LOS) biosynthesis
/ OTHER INFORMATION: locus)
/ US-09-816-028A-15

Query Match          0.3%; Score 21; DB 3; Length 1170;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6195 TATTAATGTGAAAAATATTTA 6215
DB 31 TATTAATGTGAAAAATATTTA 51

RESULT 15
US-10-303-162-15
/ Sequence 15, Application US/10303162
/ Patent No. 6723545
/ GENERAL INFORMATION:
/ APPLICANT: Gilbert, Michel
/ APPLICANT: Wakarchuk, Warren W.
/ APPLICANT: National Research Council of Canada
/ TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
/ FILE REFERENCE: 019633-00011US
/ CURRENT APPLICATION NUMBER: US/10/303,162
/ CURRENT FILING DATE: 2002-11-21
/ PRIOR APPLICATION NUMBER: US/09/816,028
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: US 60/118,213
/ PRIOR FILING DATE: 1999-02-01
```

PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 15
LENGTH: 1170
TYPE: DNA
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: glycosyltransferase from C. jejuni strain OH4384
OTHER INFORMATION: (ORF 4a of lipooligosaccharide (LOS) biosynthesis
OTHER INFORMATION: locus)
US-10-303-162-15

Query Match 0.3%; Score 21; DB 3; Length 1170;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6195 TATPATGTAGAAAAATATTTA 6215
|||
DB 31 TATPATGTAGAAAAATATTTA 51

Search completed: December 26, 2005, 05:14:04
Job time : 786 secs

THIS PAGE BLANK (USPTO)

Db 181 TAGAGTTGAAAAACATTTAACTATAGTATATATTTGCAATTTTCTTACGGCAGTAT 240
Qy 241 CATTTTGTGGGAATAATTTGCACTTTCAAGACGTGGTCCGTATTTTCACTTAA 300
Db 241 CATTTTGTGGGAATAATTTGCACTTTCAAGACGTGGTCCGTATTTTCACTTAA 300
Qy 301 TAAACTTCGTTTGTATACCTATTTTAAAGCTAATTTTAAAGAGTTAAGATGCTTTC 360
Db 301 TAAACTTCGTTTGTATACCTATTTTAAAGCTAATTTTAAAGAGTTAAGATGCTTTC 360
Qy 361 TATTTTCGACATCTATCAAAAAAAGACGATTTCTAATTTTCAAGGCTGAACGATGGGAAA 420
Db 361 TATTTTCGACATCTATCAAAAAAAGACGATTTCTAATTTTCAAGGCTGAACGATGGGAAA 420
Qy 421 ATATGCAAGTTTATTTGATCATATTAACAAATTCGAAAAATCTTGTGCAATGGTAG 480
Db 421 ATATGCAAGTTTATTTGATCATATTAACAAATTCGAAAAATCTTGTGCAATGGTAG 480
Qy 481 TTTTACGTACAGAAATAGATTAATTAATTTATCATTTACCGCTCTATTTATTTCTGGGAG 540
Db 481 TTTTACGTACAGAAATAGATTAATTAATTTATCATTTATCATTTACCGCTCTATTTATTTCTGGGAG 540
Qy 541 AAGCTATAGATTTTCAACAGGAAAGTGTGACACACGCTTTATTAATTAATCTAACAAATG 600
Db 541 AAGCTATAGATTTTCAACAGGAAAGTGTGACACACGCTTTATTAATTAATCTAACAAATG 600
Qy 601 AGTTTATGACGTAAGCAATTCGTTTCAATTTTGAAGTTGATGATGATGATGATGATG 660
Db 601 AGTTTATGACGTAAGCAATTCGTTTCAATTTTGAAGTTGATGATGATGATGATGATG 660
Qy 661 TTGATATTAATTCATTCGTTTTCATGCTGTAAGAAAACAAAAAATCCAACTGCTAGTG 720
Db 661 TTGATATTAATTCATTCGTTTTCATGCTGTAAGAAAACAAAAAATCCAACTGCTAGTG 720
Qy 721 ACCATAGCATTTTCACTTTTCCAAATTTTATAGCTATGATATATATATATATATAT 780
Db 721 ACCATAGCATTTTCACTTTTCCAAATTTTATAGCTATGATATATATATATATATATAT 780
Qy 781 GACTTTGATATATCTGAGCGGATGCGGGTAAATTTTGTGGTATGATGATGATGATG 840
Db 781 GACTTTGATATATCTGAGCGGATGCGGGTAAATTTTGTGGTATGATGATGATGATG 840
Qy 841 TGTAGTTCCAAATTTCTAGAGATGATGACCGGCTATTTTGTCTCAGAAAAGATG 900
Db 841 TGTAGTTCCAAATTTCTAGAGATGATGACCGGCTATTTTGTCTCAGAAAAGATG 900
Qy 901 GACAGATGACGATATTTTACATTTTCAAGATTTTCAATGATGATGATGATGATGATG 960
Db 901 GACAGATGACGATATTTTACATTTTCAAGATTTTCAATGATGATGATGATGATGATG 960
Qy 961 AGGCAAAAAAGACTGCTCAGCCAAAACAGATGCAAGGATGGGTATGTTTAAATGG 1020
Db 961 AGGCAAAAAAGACTGCTCAGCCAAAACAGATGCAAGGATGGGTATGTTTAAATGG 1020
Qy 1021 GAAAAAGCATCTAGATTTCTCAATTTGACATTTTATCGCAAAAAACAAAGTTTACG 1080
Db 1021 GAAAAAGCATCTAGATTTCTCAATTTGACATTTTATCGCAAAAAACAAAGTTTACG 1080
Qy 1081 AGTTACCAAGTTTATATGTTTAAATTTGGCGATATGATCTATGTTGTAACGCTCAC 1140
Db 1081 AGTTACCAAGTTTATATGTTTAAATTTGGCGATATGATCTATGTTGTAACGCTCAC 1140
Qy 1141 CTACAGTTGATTTTGAATAATATCTCTGCTCAAAAAGAGCGATGATGTTTAAAC 1200
Db 1141 CTACAGTTGATTTTGAATAATATCTCTGCTCAAAAAGAGCGATGATGTTTAAAC 1200
Qy 1201 CAGGATTTACAGCTCTCTGCGAGGTTAGTGTCTAGTAAATATCAACAGCTTTCAGCAG 1260
Db 1201 CAGGATTTACAGCTCTCTGCGAGGTTAGTGTCTAGTAAATATCAACAGCTTTCAGCAG 1260
Qy 1261 TAGTTGCGTTGACCTTAGCATATGATTAATGATGATCTGCTGATGATTTAAATTT 1320
Db 1261 TAGTTGCGTTGACCTTAGCATATGATTAATGATGATCTGCTGATGATTTAAATTT 1320

Qy 1321 TATTAAGACAGTGAAGTTGATTTGTTGAGAGAGGAACTAAGTAAAGTATATGAAAG 1380
Db 1321 TATTAAGACAGTGAAGTTGATTTGTTGAGAGAGGAACTAAGTAAAGTATATGAAAG 1380
Qy 1381 TTTTGTGGTGGTCTTCAGGGGGGACATTTGACTGACTGTATTTGTTTAAACGCTTT 1440
Db 1381 TTTTGTGGTGGTCTTCAGGGGGGACATTTGACTGACTGTATTTGTTTAAACGCTTT 1440
Qy 1441 GGAAGGAAAGAAAGCTTTTGGGTAACTTTGATTAAGAGATGCAAGATCTTTTGA 1500
Db 1441 GGAAGGAAAGAAAGCTTTTGGGTAACTTTGATTAAGAGATGCAAGATCTTTTGA 1500
Qy 1501 AGAATGAAAAATGATTCATGTTACTTTCAACAAATGCAATCTCATTAATTTAGTAA 1560
Db 1501 AGAATGAAAAATGATTCATGTTACTTTCAACAAATGCAATCTCATTAATTTAGTAA 1560
Qy 1561 AAAATGCTTTCTAGCTTTCAAAATTTTACGATGAGAAACCAAGTGTATTTATTCAT 1620
Db 1561 AAAATGCTTTCTAGCTTTCAAAATTTTACGATGAGAAACCAAGTGTATTTATTCAT 1620
Qy 1621 CTGTCGCGCGTGTGCTGCCCTTCTTTTACATCGGAAAACATTTTGGAGCAAGACGA 1680
Db 1621 CTGTCGCGCGTGTGCTGCCCTTCTTTTACATCGGAAAACATTTTGGAGCAAGACGA 1680
Qy 1681 TTTATATGAGATTTGATCGAGTTAATTAATCTAATTAACGTGAAAACCTAAGTTATC 1740
Db 1681 TTTATATGAGATTTGATCGAGTTAATTAATTAATCTAATTAACGTGAAAACCTAAGTTATC 1740
Qy 1741 CCGTAAACAATTTTATTTGTTTCAAGGGAAGAAATGAAAGGATATCTCAATCTTA 1800
Db 1741 CCGTAAACAATTTTATTTGTTTCAAGGGAAGAAATGAAAGGATATCTCAATCTTA 1800
Qy 1801 TTTACTGGGAGATTTTATTAATGATTTTGTAAACATGAGAACTCATGAACAAGTT 1860
Db 1801 TTTACTGGGAGATTTTATTAATGATTTTGTAAACATGAGAACTCATGAACAAGTT 1860
Qy 1861 TAAATGATTTGATTAAGATTTGATTTTGAAGAAATTTGAAAGTATTAACGACGAAT 1920
Db 1861 TAAATGATTTGATTAAGATTTGATTTTGAAGAAATTTGAAAGTATTAACGACGAAT 1920
Qy 1921 ATTTATTTCAACAGGATTTTCTGACATATTCAGAAATTTTCAAGATTAATTAATTTCT 1980
Db 1921 ATTTATTTCAACAGGATTTTCTGACATATTCAGAAATTTTCAAGATTAATTAATTTCT 1980
Qy 1981 CAGTTACAAAGAAATGAAACAATATATTAACAATGAGAAATGATTTTCCACGAGG 2040
Db 1981 CAGTTACAAAGAAATGAAACAATATATTAACAATGAGAAATGATTTTCCACGAGG 2040
Qy 2041 CCCGCTACTTTTATGAAATTCATTTATCCAAAGGAAAAACAATTTGTTCTTGACA 2100
Db 2041 CCCGCTACTTTTATGAAATTCATTTATCCAAAGGAAAAACAATTTGTTCTTGACA 2100
Qy 2101 AAAAAAGTATGGAACATGTAATGATCATCAAGATGAGTTTGTAAAGAAATTTTACA 2160
Db 2101 AAAAAAGTATGGAACATGTAATGATCATCAAGATGAGTTTGTAAAGAAATTTTACA 2160
Qy 2161 AGATTAATTAATTTTATTTATTAAGAAATATAGATGATTTGTTGAAGAAATTTTGAAGT 2220
Db 2161 AGATTAATTAATTTTATTTATTAAGAAATATAGATGATTTGTTGAAGAAATTTTGAAGT 2220
Qy 2221 TTCTAAGCAAACTTACCTTACATCAATTAATTAATTTTGTGAAGATTTAAACAAT 2280
Db 2221 TTCTAAGCAAACTTACCTTACATCAATTAATTAATTTTGTGAAGATTTAAACAAT 2280
Qy 2281 AGTTGAAAAATTTAAGAGATCAAGAAATGAATTAATAAAGATGCAATTTTGTAA 2340
Db 2281 AGTTGAAAAATTTAAGAGATCAAGAAATGAATTAATAAAGATGCAATTTTGTAA 2340
Qy 2341 TGGCTTATCATTAATTTTCTCAGATTTTACTGAGAGGAGATACAGATTTATCATCTCT 2400
Db 2341 TGGCTTATCATTAATTTTCTCAGATTTTACTGAGAGGAGATACAGATTTATCATCTCT 2400

QY 2401 CTCAGAGATGACACACCTAGTTCCTTCAGAAATACCTGATATATTTTAAATATTC 2460
DB 2401 CTCAGAGATGACACACCTAGTTCCTTCAGAAATACCTGATATATTTTAAATATTC 2460
QY 2461 TCAGATTTATATGTTGAATTTACAAAAGATGACAAAATATTAAGAAAATGATATAT 2520
DB 2461 TCAGATTTATATGTTGAATTTACAAAAGATGACAAAATATTAAGAAAATGATATAT 2520
QY 2521 TCAGAGATTTAATGTTACAGATTTATTCCTATATATATGACAAAATCTATGATATAT 2580
DB 2521 TCAGAGATTTAATGTTACAGATTTATTCCTATATATATGACAAAATCTATGATATAT 2580
QY 2581 ACTGTTAGAAATTTTATTAAGATGATGAGCTTTGAAATCTATTTACAAAGATTTGTT 2640
DB 2581 ACTGTTAGAAATTTTATTAAGATGATGAGCTTTGAAATCTATTTACAAAGATTTGTT 2640
QY 2641 GTTTATGATAGATTAATAAACAATGCTTACAGATTAAGATTTGTTCTAATGGGTTTG 2700
DB 2641 GTTTATGATAGATTAATAAACAATGCTTACAGATTAAGATTTGTTCTAATGGGTTTG 2700
QY 2701 CTTCCACATGATTTTGTGGCAATCTTTATGCAAAAGAAAACAGCTTATTTATTT 2760
DB 2701 CTTCCACATGATTTTGTGGCAATCTTTATGCAAAAGAAAACAGCTTATTTATTT 2760
QY 2761 AAGTATCTAAATGTCAGATGAACTATTTATACAGCAATTAAGAAAATGATGATTT 2820
DB 2761 AAGTATCTAAATGTCAGATGAACTATTTATACAGCAATTAAGAAAATGATGATTT 2820
QY 2821 TCAGATAGATTTATCTAAATATGAGAAATTTAAGATATATAAGTGAAGAAAATCAACAT 2880
DB 2821 TCAGATAGATTTATCTAAATATGAGAAATTTAAGATATATAAGTGAAGAAAATCAACAT 2880
QY 2881 TCTCCATTTGCTTTACAGATGATTTCTATTTGATGAAATGCTTAAATGCAAGAAATTTAG 2940
DB 2881 TCTCCATTTGCTTTACAGATGATTTCTATTTGATGAAATGCTTAAATGCAAGAAATTTAG 2940
QY 2941 TTTTATTTGCTAGAAAGTTAAATAAGAAATTAATCTAATTTTAAAGAAATTTATTTACT 3000
DB 2941 TTTTATTTGCTAGAAAGTTAAATAAGAAATTAATCTAATTTTAAAGAAATTTATTTACT 3000
QY 3001 AAAAAATATAATAGTATTTTGTGAGATATGATGTTTAAATTTTAAATATGATGAC 3060
DB 3001 AAAAAATATAATAGTATTTTGTGAGATATGATGTTTAAATTTTAAATATGATGAC 3060
QY 3061 CGGAATTTTATTTTAAAGTCTGCTGTTGATTTTATTTTCCAGAGCAAAAGATG 3120
DB 3061 CGGAATTTTATTTTAAAGTCTGCTGTTGATTTTATTTTCCAGAGCAAAAGATG 3120
QY 3121 TATTTTATTAATTTTAAAGTATTTTATTTTATTTTCAATTAATAATTTTGAAGAACTAGC 3180
DB 3121 TATTTTATTAATTTTAAAGTATTTTATTTTATTTTCAATTAATAATTTTGAAGAACTAGC 3180
QY 3181 TAAATTTAAATAATGAAATTTTATGTTTATTTATATGCTATATATGTTTGTTCAG 3240
DB 3181 TAAATTTAAATAATGAAATTTTATGTTTATTTATATGCTATATATGTTTGTTCAG 3240
QY 3241 TAGTCAAGATATGTTTGTGAATAATATTTGAAGATTTTGAAGATTTTACTGCTC 3300
DB 3241 TAGTCAAGATATGTTTGTGAATAATATTTGAAGATTTTGAAGATTTTACTGCTC 3300
QY 3301 CCAATATTTGATATTTGCAATATGATATTAATTTGATTCATTTAATAATATGAT 3360
DB 3301 CCAATATTTGATATTTGCAATATGATATTAATTTGATTCATTTAATAATATGAT 3360
QY 3361 AAAAAATATAATAGTATTTTGTGATTTTATTTTATTTTAAAGTATGAT 3420
DB 3361 AAAAAATATAATAGTATTTTGTGATTTTATTTTATTTTAAAGTATGAT 3420
QY 3421 TGTATATTTTCAAAATGGAAGATATTTTATTTTGAAGCACTTATATGAGCTAG 3480
DB 3421 TGTATATTTTCAAAATGGAAGATATTTTATTTTGAAGCACTTATATGAGCTAG 3480
QY 3481 ACTATCTTATTAACAGGCTCAAAAACAAGTTGGTTGCTTTATGACTATCTTACGTAA 3540

DB 3481 ACTATCTTATTAACAGGCTCAAAAACAAGTTGGTTGCTTTATGACTATCTTACGTAA 3540
QY 3541 ATACCACTACCAATTTATGTTTCAATTTCCGTTATCTTTGACCTTATTAATAATATG 3600
DB 3541 ATACCACTACCAATTTATGTTTCAATTTCCGTTATCTTTGACCTTATTAATAATATG 3600
QY 3601 AACAAATTTTCTGTTGCTGCTTTTATACCGATCTATTTTAAAGTATGAGATTTG 3660
DB 3601 AACAAATTTTCTGTTGCTGCTTTTATACCGATCTATTTTAAAGTATGAGATTTG 3660
QY 3661 GTAGTTATGCTAGCAATTTATTAATATGCTTTATGAGATATATAGTGAAAT 3720
DB 3661 GTAGTTATGCTAGCAATTTATTAATATGCTTTATGAGATATATAGTGAAAT 3720
QY 3721 TTGCTTGAATAAATAAGCAATATGATATTTGTAATCTACTTATTTTAAATACG 3780
DB 3721 TTGCTTGAATAAATAAGCAATATGATATTTGTAATCTACTTATTTTAAATACG 3780
QY 3781 AATGCTTATACATGAAATTTTGTGCTTTATATCTAGATCAATCAAGATTA 3840
DB 3781 AATGCTTATACATGAAATTTTGTGCTTTATATCTAGATCAATCAAGATTA 3840
QY 3841 GATTTATATTTATCAAGAGATTTGATTAAGATTTAAGAAACAATATTTTATTTGAT 3900
DB 3841 GATTTATATTTATCAAGAGATTTGATTAAGATTTAAGAAACAATATTTTATTTGAT 3900
QY 3901 ATGGAATATCCGAATTTTATGATTAAGGGAACCTTGGTCCGAATCTTCAAGCTATAT 3960
DB 3901 ATGGAATATCCGAATTTTATGATTAAGGGAACCTTGGTCCGAATCTTCAAGCTATAT 3960
QY 3961 CATTTTATTAATACAGGATATGTTGGTATTTTATCTAGATTTTCTTTTATG 4020
DB 3961 CATTTTATTAATACAGGATATGTTGGTATTTTATCTAGATTTTCTTTTATG 4020
QY 4021 TTTTAAATAAATAAGTATGAGTTTAAATGGGAAACAGCACTATTTTATTTACATCTAG 4080
DB 4021 TTTTAAATAAATAAGTATGAGTTTAAATGGGAAACAGCACTATTTTATTTACATCTAG 4080
QY 4081 CCAATTTTCAATATGAGAAACAATATGATCCGATTTATTTATATTAATGATCTTTT 4140
DB 4081 CCAATTTTCAATATGAGAAACAATATGATCCGATTTATTTATTAATGATCTTTT 4140
QY 4141 CTTCAATAGTATTTGGAATATATTAATTTTAAAGATTTGAGCAAAAGAT 4200
DB 4141 CTTCAATAGTATTTGGAATATATTAATTTTAAAGATTTTGAAGCAAAAGAT 4200
QY 4201 GATTTATTTCACTTATTTGATCAATTTTAAATGTCAGATTTATCTGATATAT 4260
DB 4201 GATTTATTTCACTTATTTGATCAATTTTAAATGTCAGATTTATCTGATATAT 4260
QY 4261 AACAGTATTTATTAACCAACAATATTAATTTAAGGATTTTCTGTAATGATGAA 4320
DB 4261 AACAGTATTTATTAACCAACAATATTAATTTAAGGATTTTCTGTAATGATGAA 4320
QY 4321 ACTGATGATTTGAGAAATTTGCTTAACTATTAAGAAAGATGGAAGATTAAT 4380
DB 4321 ACTGATGATTTGAGAAATTTGCTTAACTATTAAGAAAGATGGAAGATTAAT 4380
QY 4381 TACAGAAATTAATGCGGCTAGCAGATGCTGCAATTTTGGACTAGAACATGCA 4440
DB 4381 TACAGAAATTAATGCGGCTAGCAGATGCTGCAATTTTGGACTAGAACATGCA 4440
QY 4441 GGTAAATATATTTGCTTTGCTGATTTGATGATCTATATGAAAGTTCAGAG 4500
DB 4441 GGTAAATATATTTGCTTTGCTGATTTGATGATCTATATGAAAGTTCAGAG 4500
QY 4501 ATGCAATGATATATTAATGATATATGCGGATATGAGAGATGATTTTGTATG 4560
DB 4501 ATGCAATGATATATTAATGATATATGCGGATATGAGAGATGATTTTGTATG 4560
QY 4561 GACGAAACGGGTATTAACAAAGAAATAAGTATTTTCAATCTTTACGAGAG 4620

Db 4561 GACGAAAA CGGCTATACAAAGAAAAAGAAATAGTAATTTTCATGCTCTTAA CGAGAGA 4620
Qy 4621 GAGACTGTAAAGAAATTTTGT CAGGATCTAAATATAGAAATATATGTTTGGTCAAGCTT 4680
Db 4621 GAGACTGTAAAGAAATTTTGT CAGGATCTAAATATAGAAATATATGTTTGGTCAAGCTT 4680
Qy 4681 TATTCACGAGATATATAAAGATATAAATTCCAAATTAATATAGAAATATGCTGAG 4740
Db 4681 TATTCACGAGATATATAAAGATATAAATTCCAAATTAATATAGAAATATGCTGAG 4740
Qy 4741 GATTTGCTTTTAAATTTGAGAGCTTGTGAA CAATGTAAACGCTGTATCTTACTATA 4800
Db 4741 GATTTGCTTTTAAATTTGAGAGCTTGTGAA CAATGTAAACGCTGTATCTTACTATA 4800
Qy 4801 GAATATATATTAATTAATATGTCATCTGTACAGTTGCTTATTAATCAGAAATTTCTATA 4860
Db 4801 GAATATATATTAATTAATATGTCATCTGTACAGTTGCTTATTAATCAGAAATTTCTATA 4860
Qy 4861 AATTAATATGATTTAGTCA CAAGATTGAGAAATTAACCCCTTTAAGTTAAAAAGAGATT 4920
Db 4861 AATTAATATGATTTAGTCA CAAGATTGAGAAATTAACCCCTTTAAGTTAAAAAGAGATT 4920
Qy 4921 AGTCATATTTTGTATGCAAAAAGTTATTAAGAGAGTTAAATGTTTAAACAAATGTAT 4980
Db 4921 AGTCATATTTTGTATGCAAAAAGTTATTAAGAGAGTTAAATGTTTAAACAAATGTAT 4980
Qy 4981 TCAACGAGATTGTTGGATATAGTCTTGCCCAATATTAAGAGCTTATCGAAAAAGATA 5040
Db 4981 TCAACGAGATTGTTGGATATAGTCTTGCCCAATATTAAGAGCTTATCGAAAAAGATA 5040
Qy 5041 CGTAGATATCATTTATTAAGGCAAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG 5100
Db 5041 CGTAGATATCATTTATTAAGGCAAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG 5100
Qy 5101 TAATTTGATGAAATTTTCCCTTAAACTATATGTATATGTTATATAGAAATTTCAAAGCAG 5160
Db 5101 TAATTTGATGAAATTTTCCCTTAAACTATATGTATATGTTATATAGAAATTTCAAAGCAG 5160
Qy 5161 TAGAGTAAATAATGATATAAATATAGTATGTTGTCAGTTTAAATGATATAATAT 5220
Db 5161 TAGAGTAAATAATGATATAAATATAGTATGTTGTCAGTTTAAATGATATAATAT 5220
Qy 5221 TAAGTAGTGTATAGAAAGCATTTATTAATCAAAATTTATAAAATATATAGAAATTTAT 5280
Db 5221 TAAGTAGTGTATAGAAAGCATTTATTAATCAAAATTTATAAAATATATAGAAATTTAT 5280
Qy 5281 TAGATGATGGCTGTATAGATGATTCGTCTAAATATGCAAGAAATATGCGAAAAAGATA 5340
Db 5281 TAGATGATGGCTGTATAGATGATTCGTCTAAATATGCAAGAAATATGCGAAAAAGATA 5340
Qy 5341 AAAAGATAAAAATTTTTCATCAATCATGTGTAGATCAAAATGCTAGAAATCATAGAA 5400
Db 5341 AAAAGATAAAAATTTTTCATCAATCATGTGTAGATCAAAATGCTAGAAATCATAGAA 5400
Qy 5401 TAAAGCGAGTACAGCTGAATATATATGTTGTGCACTGTAGATGATGTTGTATAGTA 5460
Db 5401 TAAAGCGAGTACAGCTGAATATATATGTTGTGCACTGTAGATGATGTTGTATAGTA 5460
Qy 5461 GATTAGTAGAAAAATATATATTTATATATTAATAAAGATAGAAAGTATTTCTGTTGTT 5520
Db 5461 GATTAGTAGAAAAATATATATTTATATATTAATAAAGATAGAAAGTATTTCTGTTGTT 5520
Qy 5521 TGTACGCTACTTTTTCAGAAAAATATATAATTTTGAAGTAAATCCAAATATTTGATT 5580
Db 5521 TGTACGCTACTTTTTCAGAAAAATATATAATTTTGAAGTAAATCCAAATATTTGATT 5580
Qy 5581 TTGAAGCAATTAATACCGTGCAGAGACATGGAGAAAAAATTTTATATATTTGTATATA 5640
Db 5581 TTGAAGCAATTAATACCGTGCAGAGACATGGAGAAAAAATTTTATATATTTGTATATA 5640
Qy 5641 ATATAATTTTCTACTCTGTTTGTATACTATATAGAAAAAGTATCATTAACGATCTTT 5700
Db 5641 ATATAATTTTCTACTCTGTTTGTATACTATATAGAAAAAGTATCATTAACGATCTTT 5700

Qy 5701 TTCAAGAGAAATCAATGTTTACGAGAAAGATTATCTTTTAAATCTGCATTTATTAAGAAATA 5760
Db 5701 TTCAAGAGAAATCAATGTTTACGAGAAAGATTATCTTTTAAATCTGCATTTATTAAGAAATA 5760
Qy 5761 TAGATAGAGTATAGTATTTGACTGAAACATCTTATTTTATAGAGAGATACTAGTA 5820
Db 5761 TAGATAGAGTATAGTATTTGACTGAAACATCTTATTTTATAGAGAGATACTAGTA 5820
Qy 5821 CAGTAAATTTCTTTTAAAGAGGCTGTGTTTGCATTTGCAAAATTTGCAAAAACAGTGA 5880
Db 5821 CAGTAAATTTCTTTTAAAGAGGCTGTGTTTGCATTTGCAAAATTTGCAAAAACAGTGA 5880
Qy 5881 TAGATATGTTTAAAGCAAAATATATGCTGAGATTTTGAAGTATCAATTTGTAAAGTACTA 5940
Db 5881 TAGATATGTTTAAAGCAAAATATATGCTGAGATTTTGAAGTATCAATTTGTAAAGTACTA 5940
Qy 5941 TACGTTGCAAGTATTTATATATATACCTTACATTAATGTTTAAATACGAAACAGTCTATT 6000
Db 5941 TACGTTGCAAGTATTTATATATATACCTTACATTAATGTTTAAATACGAAACAGTCTATT 6000
Qy 6001 TTGACAAATTTTAAATTTTAAAGATCTTTATTAATAAATATTTATTTAACTGTTAAAG 6060
Db 6001 TTGACAAATTTTAAATTTTAAAGATCTTTATTAATAAATATTTATTTAACTGTTAAAG 6060
Qy 6061 TATCTAACAAAAATTTCTTGTCTAAAAATTTTGTATATAAGATTTGTTGCAACAAAGTT 6120
Db 6061 TATCTAACAAAAATTTCTTGTCTAAAAATTTTGTATATAAGATTTGTTGCAACAAAGTT 6120
Qy 6121 TTTAAAAAATATATATGCTTATATAATAGAAAGATATCATGATATCTATGTAATTTCTA 6180
Db 6121 TTTAAAAAATATATATGCTTATATAATAGAAAGATATCATGATATCTATGTAATTTCTA 6180
Qy 6181 TAATTTGACCTATATATATATAGAAAAATATTTATCTTAATATATAGATAGTATGTA 6240
Db 6181 TAATTTGACCTATATATATATAGAAAAATATTTATCTTAATATATAGATAGTATGTA 6240
Qy 6241 ATCAGACCTAACAAACATATAGAGATTTCTTCTGTGAATGACGCTAGTAAATTCGG 6300
Db 6241 ATCAGACCTAACAAACATATAGAGATTTCTTCTGTGAATGACGCTAGTAAATTCGG 6300
Qy 6301 AAGAAATTTGTTTATGATATGCGAAAGAAAGATAGTCGATTTGTTAAAAAGATA 6360
Db 6301 AAGAAATTTGTTTATGATATGCGAAAGAAAGATAGTCGATTTGTTAAAAAGATA 6360
Qy 6361 ACGGCGGCTATCAGATGCGCCGTAATATAGCATTAAGTGCAGCAAGGCTGACTAG 6420
Db 6361 ACGGCGGCTATCAGATGCGCCGTAATATATAGCATTAAGTGCAGCAAGGCTGACTAG 6420
Qy 6421 CTTTATATACCTCAGATGATTTTATTCATTCGAGATTCATCCAAAGTTTACAGAAACAA 6480
Db 6421 CTTTATATACCTCAGATGATTTTATTCATTCGAGATTCATCCAAAGTTTACAGAAACAA 6480
Qy 6481 TTGAGAGAGAAATGCCCTTGTGCGAGTGTCTGTATATAGGATAGTCTTCGGGCG 6540
Db 6481 TTGAGAGAGAAATGCCCTTGTGCGAGTGTCTGTATATAGGATAGTCTTCGGGCG 6540
Qy 6541 ATTTCTTAACAGAGAGCGCTTCCATCAAAATCAGGCTGTTCTGACGCGCAGGAATGTT 6600
Db 6541 ATTTCTTAACAGAGAGCGCTTCCATCAAAATCAGGCTGTTCTGACGCGCAGGAATGTT 6600
Qy 6601 GTAAAAAGCTGCTAGAGCGGATGTGATCTGCTTGTGTGCGCTGTAAATACTCTATA 6660
Db 6601 GTAAAAAGCTGCTAGAGCGGATGTGATCTGCTTGTGTGCGCTGTAAATACTCTATA 6660
Qy 6661 AAAAAGACTATTTGAAGATTTTCGATTTGAAAAAGGTAAGTTCATGAAGATGATACT 6720
Db 6661 AAAAAGACTATTTGAAGATTTTCGATTTGAAAAAGGTAAGTTCATGAAGATGATACT 6720
Qy 6721 TCACTTATGCTTGTCTATATGATTTGAAGAAAGTTCATATAGTAGAGTGTCTTACT 6780
Db 6721 TCACTTATGCTTGTCTATATGATTTGAAGAAAGTTCATATAGTAGAGTGTCTTACT 6780

Qy	6781	ATTATGTTGACCGAGAAATATGATATCAAACTTTAGAGATAGTCAATGCGCTTCATT	6844
Db	6781	ATTATGTTGACCGAGAAATATGATATCAAACTTTAGAGATAGTCAATGCGCTTCATT	6844
Qy	6841	GCCCTACTGGAATTTCCAAATAGAACGAATGCACTTCTATGAAAGTAGAGAGATTAAGAGC	6900
Db	6841	GCCCTACTGGAATTTCCAAATAGAACGAATGCACTTCTATGAAAGTAGAGAGATTAAGAGC	6900
Qy	6901	TCTTACTAGAGTGTATCTGTTCACTTTTACGCTTTGCTGTTTGTTTTGTAGCAATATATA	6966
Db	6901	TCTTACTAGAGTGTATCTGTTCACTTTTACGCTTTGCTGTTTGTTTTGTAGCAATATATA	6966
Qy	6961	ATCATTTGGTTGACAAACAGCAAAAGAGCTT	6992
Db	6961	ATCATTTGGTTGACAAACAGCAAAAGAGCTT	6992
RESULT 2			
US-09-767-041-29			
Sequence 29, Application US/09767041			
Patent No. US2002005168A1			
GENERAL INFORMATION:			
APPLICANT: Smith, Hilda			
TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS			
FILE REFERENCE: 2183-4726			
CURRENT FILING DATE: US/09/767.041			
PRIOR FILING DATE: 2001-01-22			
PRIOR APPLICATION NUMBER: PCT/NL99/00460			
PRIOR FILING DATE: 1999-07-19			
PRIOR APPLICATION NUMBER: EP98202465.5			
PRIOR FILING DATE: 1998-07-22			
PRIOR APPLICATION NUMBER: EP98202467.1			
PRIOR FILING DATE: 1998-07-22			
NUMBER OF SEQ ID NOS: 53			
SOFTWARE: Patentin version 3.0			
SEQ ID NO 29			
LENGTH: 6992			
TYPE: DNA			
ORGANISM: Streptococcus suis			
FEATURE:			
NAME/KEY: misc feature			
OTHER INFORMATION: CPS1			
US-09-767-041-29			
Query Match			
Best Local Similarity 100.0%; Score 6992; DB 3; Length 6992;			
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	ATCGCCAAACGAAATGGCACTTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA	60
Db	1	ATCGCCAAACGAAATGGCACTTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA	60
Qy	61	CAAGTCATATACCAAAGCGATTAATATCGTTCGGAATTTTATACATATGATAGTGTC	120
Db	61	CAAGTCATATACCAAAGCGATTAATATCGTTCGGAATTTTATACATATGATAGTGTC	120
Qy	121	ATTATTTTGCATTTTATATCTCGTATGCCAGTTGATTTAGATATAGAGTATCTGA	180
Db	121	ATTATTTTGCATTTTATATCTCGTATGCCAGTTGATTTAGATATAGAGTATCTGA	180
Qy	181	TAGAGTTGAAAAAACATTTTACTATATATATTTTGCATTTTCTTACGCGAGTAT	240
Db	181	TAGAGTTGAAAAAACATTTTACTATATATATTTTGCATTTTCTTACGCGAGTAT	240
Qy	241	CATTTTGTGGGAATTAATTTGCACTTTGCAAGCGGCGCGTATTTTCAATTA	300
Db	241	CATTTTGTGGGAATTAATTTGCACTTTGCAAGCGGCGCGTATTTTCAATTA	300
Qy	301	TAAACTTGTTTGGTATACCTATTTAAAGTATTTAATTAAGAGTTTAAGATAGCTTTC	360
Db	301	TAAACTTGTTTGGTATACCTATTTAAAGTATTTAATTAAGAGTTTAAGATAGCTTTC	360
Qy	361	TATTTTGCACATCTATCAAAAAAAGCAGTCTTATATTACAACGCGTGAACGATGGGAAA	420

D	b	361	TATTTTGGACATCTATCAAAAAAGACGATTCTAATTACACCGCTGAACGATGGAAA	420
O	y	421	ATATGCAAGTTTATTTGATACATAAACAATTCAAAAAATCTTGTGATTTGGAG	480
D	b	421	ATATGCAAGTTTATTTGAAATCAATATAACAATTCAAAAAACTTGTTGCATTTGGAG	480
O	y	481	TTTTTAGGACGAAGATGATAAATTTAATTATCAATTAACGCTCATTAATCTGTGGAG	540
D	b	481	TTTTTAGGACGAAGATGATAAATTTAATTATCAATTAACGCTCATTAATCTGTGGAG	540
O	y	541	AAGCTATAGATTTTTCAACAAGGAGAGTGTGCAACAAGTCTTTATAATCTACCAAGTG	600
D	b	541	AAGCTATAGATTTTTCAACAAGGAGAGTGTGCAACAAGTCTTTATAATCTACCAAGTG	600
O	y	601	AGTTTTTAGACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTAGATTTGATGTAAACG	660
D	b	601	AGTTTTTAGACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTAGATTTGATGTAAACG	660
O	y	661	TTGATATTAATTCATTCGGTTTTTCTCGGTTGAAAAAACAATAAATCCACTGCTAGTG	720
D	b	661	TTGATATTAATTCATTCGGTTTTTCTCGGTTGAAAAAACAATAAATCCACTGCTAGTG	720
O	y	721	ACCATAGCATTTGTAACCTTTTCCAAATTTTATAAGCTAGATCATTCATGATGAAC	780
D	b	721	ACCATAGCATTTGTAACCTTTTCCAAATTTTATAAGCTAGATCATTCATGATGAAC	780
O	y	781	GACTTTTGGATATCTCGAGACGGTAGTCGGGTTAATTAATTTGTGTATAGTTCTAATT	840
D	b	781	GACTTTTGGATATCTCGAGACGGTAGTCGGGTTAATTAATTTGTGTATAGTTCTAATT	840
O	y	841	TGTTAGTTCCATTAATTCGTAGAGATGTGTGACCGGCTAATTTTGTCTCAGAAACAAGTTG	900
D	b	841	TGTTAGTTCCATTAATTCGTAGAGATGTGTGACCGGCTAATTTTGTCTCAGAAACAAGTTG	900
O	y	901	GACAGATGACGCAATTTTCAATTCACAAAGTTTCCATCGATGATGATGTATGCTGAG	960
D	b	901	GACAGATGACGCAATTTTCAATTCACAAAGTTTCCATCGATGATGATGTATGCTGAG	960
O	y	961	AGCCCAAAAAAGACTTGCTGACGCCAAAACAGATGCAAGGGTGGGTAATGTTTAAATG	1020
D	b	961	AGCCCAAAAAAGACTTGCTGACGCCAAAACAGATGCAAGGGTGGGTAATGTTTAAATG	1020
O	y	1021	GAAAAAGCATCTTGAATTAATCTCAATTTGACATTTATACGCAAAAAACAAGTTTAGCG	1080
D	b	1021	GAAAAAGCATCTTGAATTAATCTCAATTTGACATTTATACGCAAAAAACAAGTTTAGCG	1080
O	y	1081	AGTTAACCAAGTTTATAAGTTTTTAATTTGCGCATGAGATGATGATGCTGATCAAGTCCAC	1140
D	b	1081	AGTTAACCAAGTTTATAAGTTTTTAATTTGCGCATGAGATGATGATGCTGATCAAGTCCAC	1140
O	y	1141	CTACAGTTGATGAATTTGAAAAATATATCTCGTGCTCAAAAGACAGATTGAGTTTAAAC	1200
D	b	1141	CTACAGTTGATGAATTTGAAAAATATATCTCGTGCTCAAAAGACAGATTGAGTTTAAAC	1200
O	y	1201	CAGGGATTTACAGGCTCTCGGCAAGGTTAGTGCTAGTAATATACAGACTTCCAGCAGCG	1260
D	b	1201	CAGGGATTTACAGGCTCTCGGCAAGGTTAGTGCTAGTAATATACAGACTTCCAGCAGCG	1260
O	y	1261	TAGTTCCGTTGAGACTTACATACATTTGATATTTGACATTAATCTGCTCAGATATTAAAAATTT	1320
D	b	1261	TAGTTCCGTTGAGACTTACATACATTTGATATTTGACATTAATCTGCTCAGATATTAAAAATTT	1320
O	y	1321	TATTTAAAGACAGTGAAGTTGATTTGTTGAGAGAGGGAAGTAAGTAAAGATATGAAAG	1380
D	b	1321	TATTTAAAGACAGTGAAGTTGATTTGTTGAGAGAGGGAAGTAAGTAAAGTATGAAAG	1380
O	y	1381	TTTGTGTTGCTCGCTTCTTCAAGGGGGAACATTTGACTCTGTATTTGTAAACCGTTTTT	1440
D	b	1381	TTTGTGTTGCTCGCTTCTTCAAGGGGGAACATTTGACTCTGTATTTGTAAACCGTTTTT	1440
O	y	1441	GGAGGAAAGAAAGACGTTTTTGGGTAACTTGTGTAAGAGGATGCAAGAAAGCTTTTGA	1500

Db 1441 GGAAGAGAGAGACGTTTTGGGTAACTTTGATTAAGAGATGCAAGAGCTTTTGA 1500
Qy 1501 AGAATGAAAAAATGATCCATGTTACTTTCAAACAATCGCAATCTCATTTAATTAGTA 1560
Db 1501 AGAATGAAAAAATGATCCATGTTACTTTCAAACAATCGCAATCTCATTTAATTAGTA 1560
Qy 1561 AAAAATCTTCTTACGCTTTCAAAATTTTACGTGATGAGAAACCAAGTGTATTATTCTAT 1620
Db 1561 AAAAATCTTCTTACGCTTTCAAAATTTTACGTGATGAGAAACCAAGTGTATTATTCTAT 1620
Qy 1621 CTGGTCCGCGCTGTGCTGCCCTTCTTTTACATCGGAAAACATTTTGGAGCAAAACGA 1680
Db 1621 CTGGTCCGCGCTGTGCTGCCCTTCTTTTACATCGGAAAACATTTTGGAGCAAAACGA 1680
Qy 1681 TTTATATTGAAGTATTGATCGAGTAAATCTACATTAACTGAAAACTAGTTATTC 1740
Db 1681 TTTATATTGAAGTATTGATCGAGTAAATCTACATTAACTGAAAACTAGTTATTC 1740
Qy 1741 CCGTAAACAGATATTTTATTTGTTCACTGGGAGAGAAATGAGAGGTATATCTTAAATCTA 1800
Db 1741 CCGTAAACAGATATTTTATTTGTTCACTGGGAGAGAAATGAGAGGTATATCTTAAATCTA 1800
Qy 1801 TTTACTGGGGAGATATTTTATATGATTTTGTAAACATGACATGACATGAAACAGAT 1860
Db 1801 TTTACTGGGGAGATATTTTATATGATTTTGTAAACATGACATGACATGAAACAGAT 1860
Qy 1861 TTAACCTGGGAGATATTTTATATGATTTTGTAAACATGACATGACATGAAACAGAT 1920
Db 1861 TTAACCTGGGAGATATTTTATATGATTTTGTAAACATGACATGACATGAAACAGAT 1920
Qy 1921 ATTTATTCAAACAGATATTTCTGACTATATTCAGAAATATTCGAAATATTAATAAATTTCT 1980
Db 1921 ATTTATTCAAACAGATATTTCTGACTATATTCAGAAATATTCGAAATATTAATAAATTTCT 1980
Qy 1981 CAGTTACAAAGAAATGAGAAACAATATTTAAACAATGAAAGTATTTGTCACAGAGAG 2040
Db 1981 CAGTTACAAAGAAATGAGAAACAATATTTAAACAATGAAAGTATTTGTCACAGAGAG 2040
Qy 2041 CCCCCGTACTTTTATGAAATCATTAATCCAAAGGAAAAAACAATTTTGTTCCTAGACA 2100
Db 2041 CCCCCGTACTTTTATGAAATCATTAATCCAAAGGAAAAAACAATTTTGTTCCTAGACA 2100
Qy 2101 AAAAAAGTATGAGAAATGATGATCAATCAAGTATGATTTGTAAGAAATTTTACA 2160
Db 2101 AAAAAAGTATGAGAAATGATGATCAATCAAGTATGATTTGTAAGAAATTTTACA 2160
Qy 2161 AGATAATTAATTTTATTTATAGAAAATATAGATGATTTGTTGAAAAATTTATGAGT 2220
Db 2161 AGATAATTAATTTTATTTATAGAAAATATAGATGATTTGTTGAAAAATTTATGAGT 2220
Qy 2221 TTTCTAGCAAACTAACCTTTACATCAAAATATATTTTGTGAAAGATTAAACAAT 2280
Db 2221 TTTCTAGCAAACTAACCTTTACATCAAAATATATTTTGTGAAAGATTAAACAAT 2280
Qy 2281 AGTTGAAAAATTTAATGAGATCAAGAAATGATATATATATATATATATATATATATAT 2340
Db 2281 AGTTGAAAAATTTAATGAGATCAAGAAATGATATATATATATATATATATATATATAT 2340
Qy 2341 TGGCTTATCATATTTTCTCAGATTTTACTGAGAGGAGATACAGATATATATATATATAT 2400
Db 2341 TGGCTTATCATATTTTCTCAGATTTTACTGAGAGGAGATACAGATATATATATATATAT 2400
Qy 2401 CTCAGAGAAATGACCACTAGTATTCCTGAGAAATCTGTATATATATATATATATATAT 2460
Db 2401 CTCAGAGAAATGACCACTAGTATTCCTGAGAAATCTGTATATATATATATATATATAT 2460
Qy 2461 TCAGATTTTATGATTTTACAAAAGATGAGCAAAATATATATATATATATATATATATAT 2520
Db 2461 TCAGATTTTATATGATTTTACAAAAGATGAGCAAAATATATATATATATATATATATAT 2520
Qy 2521 TGAACGAGTTAAATGTTACAGATTTATCTTAAATATATCAGAAAAAATCTATGATATATG 2580
Db 2521 TGAACGAGTTAAATGTTACAGATTTATCTTAAATATATCAGAAAAAATCTATGATATATG 2580

Qy 2581 ACTGTTTGAATTTTATTAAGATGATGACCTTTTGAATATCTATTAATAAAGATTTGT 2640
Db 2581 ACTGTTTGAATTTTATTAAGATGATGACCTTTTGAATATCTATTAATAAAGATTTGT 2640
Qy 2641 GTTTTGTGATGAAT 2700
Db 2641 GTTTTGTGATGAAT 2700
Qy 2701 CTTCACATGATTTTGTGCAATTTCTTTATCAAAATGAAAAAGAAACAGCTTATTTATTT 2760
Db 2701 CTTCACATGATTTTGTGCAATTTCTTTATCAAAATGAAAAAGAAACAGCTTATTTATTT 2760
Qy 2761 AAGTAACTAAATGTCAGATGATCTATTTATACAGCAATTTATGAAAAATATGAAATTT 2820
Db 2761 AAGTAACTAAATGTCAGATGATCTATTTATACAGCAATTTATGAAAAATATGAAATTT 2820
Qy 2821 TCAAAATGATTTCTAAATATGCAATTTAATATATATATATATATATATATATATATATAT 2880
Db 2821 TCAAAATGATTTCTAAATATGCAATTTAATATATATATATATATATATATATATATATAT 2880
Qy 2881 TCTCCATTTGCTTTACAGATGATCTATTTGATGATGCTAAATGCAAGAAATTTAGGT 2940
Db 2881 TCTCCATTTGCTTTACAGATGATCTATTTGATGATGCTAAATGCAAGAAATTTAGGT 2940
Qy 2941 TTTTATTTTGTAGAAAGTTAAATATGAAAAATTAATCTAAATTTTAAAGAAATTTATCT 3000
Db 2941 TTTTATTTTGTAGAAAGTTAAATATGAAAAATTAATCTAAATTTTAAAGAAATTTATCT 3000
Qy 3001 AAAAAATTAATATGTTGATTTTGTGAGATATATGATGTTTAAATTTAATAATATGAC 3060
Db 3001 AAAAAATTAATATGTTGATTTTGTGAGATATATGATGTTTAAATTTAATAATATGAC 3060
Qy 3061 CCGAATATTTTATTTTAAAGTACTTCTGGTGAATTTTATTTTATTCAGAGCAAAAGTATG 3120
Db 3061 CCGAATATTTTATTTTAAAGTACTTCTGGTGAATTTTATTTTATTCAGAGCAAAAGTATG 3120
Qy 3121 TATTTTATTAATTTTATGAAATTTAATTTATTCATATATATATTTTGAATACTAGC 3180
Db 3121 TATTTTATTAATTTTATGAAATTTAATTTATTCATATATATATTTTGAATACTAGC 3180
Qy 3181 TAAATATTAATAAATGAAATTTTATGTTTATATATATATATATATATATATATATATAT 3240
Db 3181 TAAATATTAATAAATGAAATTTTATGTTTATATATATATATATATATATATATATATAT 3240
Qy 3241 TAGTCAAGATATGTTTGTGAAATTTTGAAGATTTATTTTGAAGATTTTACGCTC 3300
Db 3241 TAGTCAAGATATGTTTGTGAAATTTTGAAGATTTATTTTGAAGATTTTACGCTC 3300
Qy 3301 CCAATATTTGATTTATGCAATATGATATATATATATATATATATATATATATATATAT 3360
Db 3301 CCAATATTTGATTTATGCAATATGATATATATATATATATATATATATATATATATAT 3360
Qy 3361 ATTAATAAATTTAATAAATGATCTTTTATGTTTATAGTTTATTAAGTATATCTGAT 3420
Db 3361 ATTAATAAATTTAATAAATGATCTTTTATGTTTATAGTTTATTAAGTATATCTGAT 3420
Qy 3421 TGTATATTTATCAAAATGAGAAAGATGATTTTATAGACAGACCTTATAGACATG 3480
Db 3421 TGTATATTTATCAAAATGAGAAAGATGATTTTATAGACAGACCTTATAGACATG 3480
Qy 3481 ACTATCTTATTAACAGGCGTCAAAACAGGTTGGCTTTATGAACTATCTTAA 3540
Db 3481 ACTATCTTATTAACAGGCGTCAAAACAGGTTGGCTTTATGAACTATCTTAA 3540
Qy 3541 ATACCACTAACATTAATGATTTTCAATTCGTTATCTTTGACATTTAATAAATATGAC 3600
Db 3541 ATACCACTAACATTAATGATTTTCAATTCGTTATCTTTGACATTTAATAAATATGAC 3600
Qy 3601 AACAAATTTTCTTGTGCTGCTTGTATACCGATCTATTTAAGTATGAGATGAGATTTG 3660
Db 3601 AACAAATTTTCTTGTGCTGCTTGTATACCGATCTATTTAAGTATGAGATGAGATTTG 3660

QY 3661 GAGATTATGCTAGCAATATTAATTAATGCTTTGATGAGATATATAGTGGAAAT 3720
| | | | |
Db 3661 GTAGTTATCGTAGCAATATTAATTAATGCTTTGATGAGATATATAGTGGAAAT 3720
| | | | |
QY 3721 TTGCTTGATAAAAAGCTAATGTATATTTGTAATACTATTTATTTTAATATCTG 3780
| | | | |
Db 3721 TTGCTTGATAAAAAGCTAATGTATATTTGTAATACTATTTATTTTAATATCTG 3780
| | | | |
QY 3781 AATGCTTTACAGTAATTTTGGCTGTTATATATTTAGAGATCAAGTACGAGCTA 3840
| | | | |
Db 3781 AATGCTTTACAGTAATTTTGGCTGTTATATATTTAGAGATCAAGTACGAGCTA 3840
| | | | |
QY 3841 GATTTATTTATTCAGAGAGATTCGATTAAGATTTAGAAAACAATATTTTATTTGGAT 3900
| | | | |
Db 3841 GATTTATTTATTCAGAGAGATTCGATTAAGATTTAGAAAACAATATTTTATTTGGAT 3900
| | | | |
QY 3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAGTCATTCAAGCTATATAT 3960
| | | | |
Db 3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTGGCTCGGAGTCATTCAAGCTATATAT 3960
| | | | |
QY 3961 CATTTTTTTATTAATCAGAAATGTTGGGTGATTTTACTGATGTTTCTTTTTTTATG 4020
| | | | |
Db 3961 CATTTTTTTATTAATCAGAAATGTTGGGTGATTTTACTGATGTTTCTTTTTTTATG 4020
| | | | |
QY 4021 TTATTAATAAAAGTTATGAGTTAATGGGAAAACAGACTATTTTATTTACATCTAG 4080
| | | | |
Db 4021 TTATTAATAAAAGTTATGAGTTAATGGGAAAACAGACTATTTTATTTACATCTAG 4080
| | | | |
QY 4081 CCATATTTTTGATATATGAAAACATATGATCCGATTTATTTATTTATTTAGTACTATCTTTT 4140
| | | | |
Db 4081 CCATATTTTTGATATATGAAAACATATGATCCGATTTATTTATTTATTTAGTACTATCTTTT 4140
| | | | |
QY 4141 CTTTCATAGTATTTGGAAATATTAATTTTAAAAAGATATGAGACAAAAATGAT 4200
| | | | |
Db 4141 CTTTCATAGTATTTGGAAATATTAATTTTAAAAAGATATGAGACAAAAATGAT 4200
| | | | |
QY 4201 GATTTAATTTCACTTATGTAACCAATTTATATGTCGAAGATATCTTGATTAATGAT 4260
| | | | |
Db 4201 GATTTAATTTCACTTATGTAACCAATTTATATGTCGAAGATATCTTGATTAATGAT 4260
| | | | |
QY 4261 AACAGTATTTAATACCAACATATACTAATTTTGAAGTTATCTCGTAAATGATGAAAT 4320
| | | | |
Db 4261 AACAGTATTTAATACCAACATATACTAATTTTGAAGTTATCTCGTAAATGATGAAAT 4320
| | | | |
QY 4321 ACTGATGATTTCTGAGAAAATTTTGTCTAACTATATGAGAAACGATGGAATTAATAT 4380
| | | | |
Db 4321 ACTGATGATTTCTGAGAAAATTTTGTCTAACTATATGAGAAACGATGGAATTAATAT 4380
| | | | |
QY 4381 TACAAAGAAATTAATGCGGCTCTAGCAGATGTCGAAATTTGGAATCTAGAAACATGCAACA 4440
| | | | |
Db 4381 TACAAAGAAATTAATGCGGCTCTAGCAGATGTCGAAATTTGGAATCTAGAAACATGCAACA 4440
| | | | |
QY 4441 GGTAAATATATGCTTTTGTGATTCGATTCGATATATAGAGTTGCAATGTTGAGAGA 4500
| | | | |
Db 4441 GGTAAATATATGCTTTTGTGATTCGATTCGATATATAGAGTTGCAATGTTGAGAGA 4500
| | | | |
QY 4501 ATGCAATGATATATTAATCTGAGATTAATGCCGATATAGCAGAGATAGATTTTGTGTA 4560
| | | | |
Db 4501 ATGCAATGATATATTAATCTGAGATTAATGCCGATATAGCAGAGATAGATTTTGTGTA 4560
| | | | |
QY 4561 GACGAAAACGGGATTAACAAAGAAAAAGAAATAGTAATTTTCATGCTTAAGAGAGAA 4620
| | | | |
Db 4561 GACGAAAACGGGATTAACAAAGAAAAAGAAATAGTAATTTTCATGCTTAAGAGAGAA 4620
| | | | |
QY 4621 GAGACTGTAAAGAAATTTTGTGAGATCTAATATAGAAAAATATGTTGGTCAAGCTT 4680
| | | | |
Db 4621 GAGACTGTAAAGAAATTTTGTGAGATCTAATATAGAAAAATATGTTGGTCAAGCTT 4680
| | | | |
QY 4681 TATTCACGAGATTTTATTAAGATATTAATTTCCAAATTAATTAAGAGATTTGGTGAG 4740
| | | | |
Db 4681 TATTCACGAGATTTTATTAAGATATTAATTTCCAAATTAATTAAGAGATTTGGTGAG 4740
| | | | |
QY 4741 GATTTGCTTTTATTAATTTGAGAGTCTTGAAACATGTAAACGCTGATGATGATCTAG 4800
| | | | |

Db 4741 GATTTGCTTTTATTAATTTGAGAGTCTTGAAACATGTAAACGCTGATGATGATCTAG 4800
| | | | |
QY 4801 GAATATTAATTAATTAATGCTATTCGTAACAGTTGCTTATTAATCAAGAAATCTCTATA 4860
| | | | |
Db 4801 GAATATTAATTAATTAATGCTATTCGTAACAGTTGCTTATTAATCAAGAAATCTCTATA 4860
| | | | |
QY 4861 AATATATTAATTAATTAATGCTATTCGTAACAGTTGCTTATTAATCAAGAAATCTCTATA 4920
| | | | |
Db 4861 AATATATTAATTAATTAATGCTATTCGTAACAGTTGCTTATTAATCAAGAAATCTCTATA 4920
| | | | |
QY 4921 AGTCATTAATTTGATGCAAAAAGTTAATTAAGAGAGTTAATGTTTAAACAAATGTAT 4980
| | | | |
Db 4921 AGTCATTAATTTGATGCAAAAAGTTAATTAAGAGAGTTAATGTTTAAACAAATGTAT 4980
| | | | |
QY 4981 TCAACAGATTTGTTGATTAATGAGTTCTTGCCCAATATTAAGCTTTTACGAAAAGATA 5040
| | | | |
Db 4981 TCAACAGATTTGTTGATTAATGAGTTCTTGCCCAATATTAAGCTTTTACGAAAAGATA 5040
| | | | |
QY 5041 CGTAGATATCCATTTATTAAGGAAAAGATTTTATCAAGAAACATTTAGTACGTTG 5100
| | | | |
Db 5041 CGTAGATATCCATTTATTAAGGAAAAGATTTTATCAAGAAACATTTAGTACGTTG 5100
| | | | |
QY 5101 TATTTGATGAATTTTGCCTTAACTATATGTAATGTAATTAAGAAATTTCAAAAGCAG 5160
| | | | |
Db 5101 TATTTGATGAATTTTGCCTTAACTATATGTAATGTAATTAAGAAATTTCAAAAGCAG 5160
| | | | |
QY 5161 TAGAGTAAATAATGATTAATAATTAAGTTATGTTCCAGTTTATATGTAATTAAT 5220
| | | | |
Db 5161 TAGAGTAAATAATGATTAATAATTAAGTTATGTTCCAGTTTATATGTAATTAAT 5220
| | | | |
QY 5221 TAAGTATGTTATAGAAAGCATTAATCAAAATTAATAAAATATAGAAATTTATGTA 5280
| | | | |
Db 5221 TAAGTATGTTATAGAAAGCATTAATCAAAATTAATAAAATATAGAAATTTATGTA 5280
| | | | |
QY 5281 TAGATGATGCTCTGATAGATATTCGCTAAATATGCAAGATATGCAAAAAGATA 5340
| | | | |
Db 5281 TAGATGATGCTCTGATAGATATTCGCTAAATATGCAAGATATGCAAAAAGATA 5340
| | | | |
QY 5341 AAAAGTAAATAATTTTTCACATATCATATGCGAGATCAATATGCTAGAAATCATGGA 5400
| | | | |
Db 5341 AAAAGTAAATAATTTTTCACATATCATATGCGAGATCAATATGCTAGAAATCATGGA 5400
| | | | |
QY 5401 TAAAGGAGGATACAGCGAATATTAATGTTGTGCTGATGATGTTGATGTA 5460
| | | | |
Db 5401 TAAAGGAGGATACAGCGAATATTAATGTTGTGCTGATGATGTTGATGTA 5460
| | | | |
QY 5461 GATTTAGTAAATAATTAATTTTAAATTAATAAAAGTAAAGTATTTACTGTTGT 5520
| | | | |
Db 5461 GATTTAGTAAATAATTAATTTTAAATTAATAAAAGTAAAGTATTTACTGTTGT 5520
| | | | |
QY 5521 TGTAGCTACTTTTTCAGAAAATATTAATTTTGAAGTGAATATCCAAATATGAT 5580
| | | | |
Db 5521 TGTAGCTACTTTTTCAGAAAATATTAATTTTGAAGTGAATATCCAAATATGAT 5580
| | | | |
QY 5581 TTGAACCAATTAATACCGGACAGACATGGGAGAAAAATTTTATGAATTTGTATATTA 5640
| | | | |
Db 5581 TTGAACCAATTAATTAATACCGGACAGACATGGGAGAAAAATTTTATGAATTTGTATATTA 5640
| | | | |
QY 5641 ATATATTTTCTTACTCCGTTTGTAACTATATTAAGAAAGATATCAATACAGATCTT 5700
| | | | |
Db 5641 ATATATTTTCTTACTCCGTTTGTAACTATATTAAGAAAGATATCAATACAGATCTT 5700
| | | | |
QY 5701 TTCAAGAGAAATCAATGTTTGAAGAAAGATTTACTTTTATTCGATTTATTAAGATA 5760
| | | | |
Db 5701 TTCAAGAGAAATCAATGTTTGAAGAAAGATTTACTTTTATTCGATTTATTAAGATA 5760
| | | | |
QY 5761 TAGATGAGTTATGTTATTTGATGAAATCTTATTTTATTAAGAGGATATCTAAGTA 5820
| | | | |
Db 5761 TAGATGAGTTATGTTATTTGATGAAATCTTATTTTATTAAGAGGATATCTAAGTA 5820
| | | | |
QY 5821 CAGTAAATCTTTTAAAGAGGTGTGTTTGTGCAATTTGCAAAAATTTGCAAAAGCAGTA 5880
| | | | |

```

Db      5821 CAGTAATTCCTTTAAAGAGAGTGTTTTCGAATGGAAATTTGCAAAAACAAGTGA 5880
Qy      5881 TAGTATTTGTTAAGCAATATATGTTGAGGATTTTGAAGTATCATTTGTTAAGTACTA 5940
Db      5881 TAGTATTTGTTAAGCAATATATGTTGAGGATTTTGAAGTATCATTTGTTAAGTACTA 5940
Qy      5941 TACGTTGGCAAGTATTTTATATAGCTTACTATGTTTAAATACGAAAAACAGTCTATT 6000
Db      5941 TACGTTGGCAAGTATTTTATATAGCTTACTATGTTTAAATACGAAAAACAGTCTATT 6000
Qy      6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTATPAAAAAATATATTTTAACTTTGAAG 6060
Db      6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTATPAAAAAATATATTTTAACTTTGAAG 6060
Qy      6061 TATCTAACAAAATTTCTTGTCTTAAATTTTGTATAGAAATGTTTGAACAAAGTT 6120
Db      6061 TATCTAACAAAATTTCTTGTCTTAAATTTTGTATAGAAATGTTTGAACAAAGTT 6120
Qy      6121 TTAATAAAATATATATGTTATATAAGAAAGATATCATGATCTATTAGTAAATTTCTA 6180
Db      6121 TTAATAAAATATATATGTTATATAAGAAAGATATCATGATCTATTAGTAAATTTCTA 6180
Qy      6181 TAATTGTACCTATATATATATATATATATATATATATATATATATATATATATATAT 6240
Db      6181 TAATTGTACCTATATATATATATATATATATATATATATATATATATATATATATAT 6240
Qy      6241 ATCAGACCTACAAACATATAGAAATCTTCTGTGTGAATACGGTATAGGATTAATTCG 6300
Db      6241 ATCAGACCTACAAACATATAGAAATCTTCTGTGTGAATACGGTATAGGATTAATTCG 6300
Qy      6301 AAGAAATTTGTTAGCATATGCGAAGAAAGATAGTGCATTCGTTATTTTAAAAAGAGA 6360
Db      6301 AAGAAATTTGTTAGCATATGCGAAGAAAGATAGTGCATTCGTTATTTTAAAAAGAGA 6360
Qy      6361 ACGGCGGGCTATCAGATGCCCGTAAATATATGCGAATAGCGCGCAAGGTTACTTTAG 6420
Db      6361 ACGGCGGGCTATCAGATGCCCGTAAATATATGCGAATAGCGCGCAAGGTTACTTTAG 6420
Qy      6421 CTTTATATAGCTCAGATGATTTTATTCATTCGGAAGTTTCAACGTTTACAGAAAGCA 6480
Db      6421 CTTTATATAGCTCAGATGATTTTATTCATTCGGAAGTTTCAACGTTTACAGAAAGCA 6480
Qy      6481 TTGAGAGAGAGAAATGCCCTGTGAGCAGTTGCTGTTATGATAGGATGATCTTCGAGGC 6540
Db      6481 TTGAGAGAGAGAAATGCCCTGTGAGCAGTTGCTGTTATGATAGGATGATCTTCGAGGC 6540
Qy      6541 ATTTCTTAAACAGAGAGCCGCTTCTTCAAAATCAGGCTGTCTGAGCGCAGAAATGTTT 6600
Db      6541 ATTTCTTAAACAGAGAGCCGCTTCTTCAAAATCAGGCTGTCTGAGCGCAGAAATGTTT 6600
Qy      6601 GTAAAAAGCTCTAGAGGCGGATGCTATCGCTTTGTGTGGCTGTATATAAATCTCTATA 6660
Db      6601 GTAAAAAGCTCTAGAGGCGGATGCTATCGCTTTGTGTGGCTGTATATAAATCTCTATA 6660
Qy      6661 AAAAAAGAACTATTTGAAGTTTTCGATTTGAAAAAGGTTAAGATTATCATGAGATGAAATCT 6720
Db      6661 AAAAAAGAACTATTTGAAGTTTTCGATTTGAAAAAGGTTAAGATTATCATGAGATGAAATCT 6720
Qy      6721 TCACTTATCGCTTCTATGATGATTAGAAAAAGTTGCAATAGTTAAGAGTCTCTTACT 6780
Db      6721 TCACTTATCGCTTCTATGATGATTAGAAAAAGTTGCAATAGTTAAGAGTCTCTTACT 6780
Qy      6781 ATTATGTTGACCGAAGAAATATGATCACACTTTCAGATGACGACCATCGCTTCATT 6840
Db      6781 ATTATGTTGACCGAAGAAATATGATCACACTTTCAGATGACGACCATCGCTTCATT 6840
Qy      6841 GCCTACGGAATTTCAAAATGAAGAAATGACCTTCTATGAAGAGTAAAGATTAAGAGC 6900
Db      6841 GCCTACGGAATTTCAAAATGAAGAAATGACCTTCTATGAAGAGTAAAGATTAAGAGC 6900
Qy      6901 TCTTACTAGAGTGTATCGTTCAATTTTATGAGCTTTGCTGTTTGTATTTAGCAAAATATA 6960
Db      6901 TCTTACTAGAGTGTATCGTTCAATTTTATGAGCTTTGCTGTTTGTATTTAGCAAAATATA 6960

```

```

Qy      6961 ATCATTTGTTGAGCAACAGCAAAAGAAAGCTT 6992
Db      6961 ATCATTTGTTGAGCAACAGCAAAAGAAAGCTT 6992

RESULT 3
US-10-461-990-4
; Sequence 4, Application US/10461990
; Publication No. US20040023361A1
; GENERAL INFORMATION:
; APPLICANT: Nestec, S.A.
; APPLICANT: Neeser, Jean-Richard
; APPLICANT: Mollete, B.
; APPLICANT: Singele, Francesca
; APPLICANT: Zinc, Robert
; APPLICANT: Kracky, Zoe
; TITLE OF INVENTION: Lactic Acid Bacteria Producing Polysaccharide Similar to those in
; FILE REFERENCE: 88265-10322
; CURRENT APPLICATION NUMBER: US/10/461,990
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: US 09/548,606
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: PCT/EP 98 06636
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: CH 94 203245.2
; PRIOR FILING DATE: 2000-01-01
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 18373
; TYPE: DNA
; ORGANISM: Streptococcus macedonicus
US-10-461-990-4

Query Match      0.4%; Score 27; DB 7; Length 18373;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1807 TGGGAGTATTTTAAAGATTTTGT 1833
Db      7566 TGGGAGTATTTTAAATGATTTTGT 7592

RESULT 4
US-10-804-408-162
; Sequence 162, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 162
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-162

Query Match      0.4%; Score 26; DB 8; Length 2217;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1755 TTTATTTGTTCAGTGGAGAAATGAA 1780

```

Db 1932 TTTATTGTTCAAGTGGAGAAATGAA 1957
|||||
RESULT 5
US-10-804-408-163
; Sequence 163, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 163
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-163

Query Match 0.4%; Score 26; DB 8; Length 2217;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
Db 1932 TTTATTGTTCAAGTGGAGAAATGAA 1957
|||||

RESULT 6
US-10-804-408-164
; Sequence 164, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-164

Query Match 0.4%; Score 26; DB 8; Length 2217;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
Db 1932 TTTATTGTTCAAGTGGAGAAATGAA 1957
|||||

RESULT 7
US-10-804-408-172
; Sequence 172, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:

; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 172
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-172

Query Match 0.4%; Score 26; DB 8; Length 2217;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
Db 1932 TTTATTGTTCAAGTGGAGAAATGAA 1957
|||||

RESULT 8
US-10-804-408-165
; Sequence 165, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 165
; LENGTH: 2225
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-165

Query Match 0.4%; Score 26; DB 8; Length 2225;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
Db 1940 TTTATTGTTCAAGTGGAGAAATGAA 1965
|||||

RESULT 9
US-10-804-408-166
; Sequence 166, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749

PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: PatentIn version 3.2
SEQ ID NO 166
LENGTH: 2226
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-10-804-408-166

Query Match 0.4%; Score 26; DB 8; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTCAGTGGGAAGAAATGAA 1780
DB 1941 TTTATTGTCAGTGGGAAGAAATGAA 1966

RESULT 10
US-10-804-408-167
Sequence 167, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: PatentIn version 3.2
SEQ ID NO 167
LENGTH: 2226
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-10-804-408-167

Query Match 0.4%; Score 26; DB 8; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTCAGTGGGAAGAAATGAA 1780
DB 1941 TTTATTGTCAGTGGGAAGAAATGAA 1966

RESULT 11
US-10-804-408-168
Sequence 168, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: PatentIn version 3.2
SEQ ID NO 168
LENGTH: 2226
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-10-804-408-168

Query Match 0.4%; Score 26; DB 8; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTCAGTGGGAAGAAATGAA 1780
DB 1941 TTTATTGTCAGTGGGAAGAAATGAA 1966

RESULT 12
US-10-804-408-169
Sequence 169, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: PatentIn version 3.2
SEQ ID NO 169
LENGTH: 2226
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-10-804-408-169

Query Match 0.4%; Score 26; DB 8; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTCAGTGGGAAGAAATGAA 1780
DB 1941 TTTATTGTCAGTGGGAAGAAATGAA 1966

RESULT 13
US-10-804-408-170
Sequence 170, Application US/10804408
Publication No. US20040253617A1
GENERAL INFORMATION:
APPLICANT: FANRONG, KONG
APPLICANT: GILBERT, GWENDOLYN
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
FILE REFERENCE: 675002-2001
CURRENT APPLICATION NUMBER: US/10/804,408
CURRENT FILING DATE: 2004-03-19
PRIOR APPLICATION NUMBER: PCT/AU02/01281
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: AU PR 7749
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 182
SOFTWARE: PatentIn version 3.2
SEQ ID NO 170
LENGTH: 2226
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-10-804-408-170

Query Match 0.4%; Score 26; DB 8; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTCAGTGGGAAGAAATGAA 1780
DB 1941 TTTATTGTCAGTGGGAAGAAATGAA 1966

RESULT 14
US-10-804-408-171
; Sequence 171, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 171
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-171

Query Match 0.4%; Score 26; DB 8; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
DB 1941 TTTATTGTTCAAGTGGAGAAATGAA 1966

RESULT 15
US-10-804-408-173
; Sequence 173, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 173
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-804-408-173

Query Match 0.4%; Score 26; DB 8; Length 2226;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 TTTATTGTTCAAGTGGAGAAATGAA 1780
DB 1941 TTTATTGTTCAAGTGGAGAAATGAA 1966

Search completed: December 26, 2005, 06:09:32
Job time : 3320 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2005, 18:09:26 ; Search time 609 Seconds

(without alignments)
5958.181 Million cell updates/sec

Title: US-09-767-041-9

Perfect score: 6992

Sequence: 1 atcgccaacgaatgcga.....gcaaacgcaaaagagctt 6992

Scoring table: OLIGO_NUC

Gapop 60.0 ; Gapext 60.0

Searched: 4168288 seqs, 259477437 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA New:
1: /cgn2_6/ptoddata/1/pubpna/US08_NEW_PUB_seq:*
2: /cgn2_6/ptoddata/1/pubpna/US06_NEW_PUB_seq:*
3: /cgn2_6/ptoddata/1/pubpna/US07_NEW_PUB_seq:*
4: /cgn2_6/ptoddata/1/pubpna/PCT_NEW_PUB_seq:*
5: /cgn2_6/ptoddata/1/pubpna/US09_NEW_PUB_seq:*
6: /cgn2_6/ptoddata/1/pubpna/US10_NEW_PUB_seq:*
7: /cgn2_6/ptoddata/1/pubpna/US11_NEW_PUB_seq:*
8: /cgn2_6/ptoddata/1/pubpna/US11_NEW_PUB_seq:*
9: /cgn2_6/ptoddata/1/pubpna/US11_NEW_PUB_seq:*
10: /cgn2_6/ptoddata/1/pubpna/US60_NEW_PUB_seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	0.3	1301	6 US-10-750-185-42158	Sequence 42158, A
C 2	21	0.3	2704	6 US-10-750-185-59406	Sequence 59406, A
C 3	21	0.3	146656	7 US-11-121-086-68	Sequence 68, App1
C 4	21	0.3	160213	7 US-11-121-086-103	Sequence 103, App1
C 5	20	0.3	991	6 US-10-750-185-28042	Sequence 28042, A
C 6	20	0.3	1077	6 US-10-750-185-25079	Sequence 25079, A
C 7	20	0.3	1206	6 US-10-750-185-42078	Sequence 42078, A
C 8	20	0.3	1324	6 US-10-750-185-62576	Sequence 62576, A
C 9	20	0.3	1362	7 US-11-074-176-141	Sequence 141, App
C 10	20	0.3	1390	6 US-10-750-185-56520	Sequence 56520, A
C 11	20	0.3	1826	6 US-10-750-185-35712	Sequence 35712, A
C 12	20	0.3	193363	7 US-11-112-908-32	Sequence 32, App1
C 13	20	0.3	1080000	6 US-10-928-446A-1	Sequence 1, App1
C 14	20	0.3	1080000	6 US-10-928-446A-181	Sequence 181, App
C 15	20	0.3	1080000	6 US-10-928-446A-183	Sequence 183, App
C 16	20	0.3	1080000	6 US-10-928-446A-185	Sequence 185, App
C 17	20	0.3	1080000	6 US-10-928-446A-187	Sequence 187, App
C 18	20	0.3	1080000	6 US-10-928-446A-189	Sequence 189, App
C 19	20	0.3	1080000	6 US-10-928-446A-191	Sequence 191, App
C 20	20	0.3	1080000	6 US-10-928-446A-193	Sequence 193, App
C 21	20	0.3	1080000	6 US-10-928-446A-195	Sequence 195, App
C 22	20	0.3	1080000	6 US-10-928-446A-197	Sequence 197, App
C 23	20	0.3	1080000	6 US-10-928-446A-199	Sequence 199, App

C 24	20	0.3	1080000	6 US-10-928-446A-201	Sequence 201, App
C 25	19	0.3	201	6 US-10-995-561-78775	Sequence 78775, A
C 26	19	0.3	600	6 US-10-750-185-3858	Sequence 3858, App
C 27	19	0.3	681	6 US-10-750-185-48046	Sequence 48046, A
C 28	19	0.3	707	6 US-10-750-185-49075	Sequence 49075, A
C 29	19	0.3	720	6 US-10-750-185-48171	Sequence 48171, A
C 30	19	0.3	754	6 US-10-750-185-63338	Sequence 63338, A
C 31	19	0.3	769	6 US-10-750-185-64133	Sequence 64133, A
C 32	19	0.3	795	6 US-10-750-185-30091	Sequence 30091, A
C 33	19	0.3	834	6 US-10-667-295-76	Sequence 76, App1
C 34	19	0.3	947	6 US-10-750-185-56215	Sequence 56215, A
C 35	19	0.3	990	6 US-10-750-185-54762	Sequence 54762, A
C 36	19	0.3	1034	6 US-10-750-185-61634	Sequence 61634, A
C 37	19	0.3	1101	6 US-10-750-185-53014	Sequence 53014, A
C 38	19	0.3	1249	6 US-10-750-185-52148	Sequence 52148, A
C 39	19	0.3	1283	6 US-10-750-185-54037	Sequence 54037, A
C 40	19	0.3	1299	6 US-10-750-185-28532	Sequence 28532, A
C 41	19	0.3	1576	6 US-10-750-185-39921	Sequence 39921, A
C 42	19	0.3	1630	6 US-10-750-185-49204	Sequence 49204, A
C 43	19	0.3	1639	6 US-10-750-185-40765	Sequence 40765, A
C 44	19	0.3	1970	6 US-10-750-185-41249	Sequence 41249, A
C 45	19	0.3	2148	6 US-10-750-185-57211	Sequence 57211, A

ALIGNMENTS

```

RESULT 1
US-10-750-185-42158/c
/ Sequence 42158, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMT GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 42158
/ LENGTH: 1301
/ TYPE: DNA
/ ORGANISM: Bovine 19866881137488
US-10-750-185-42158

Query Match 0.3%; Score 22; DB 6; Length 1301;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2721 AATCTTTTATCAATGAAAC 2742
Db 419 AATCTTTTATCAATGAAAC 398

RESULT 2
US-10-750-185-59406
/ Sequence 59406, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMT GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen

```

```

; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59406
; LENGTH: 2704
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-59406

Query Match          0.3%; Score 21; DB 6; Length 2704;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3997 TACTGATGTTTCTTTT 4017
Db      582 TACTGATGTTTCTTTT 602

RESULT 3
; Sequence 68, Application US/1121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR FILING DATE: 2004-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 68
; LENGTH: 146656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-68

Query Match          0.3%; Score 21; DB 7; Length 146656;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3115 AGTATGATTTTATTAATT 3135
Db      91893 AGTATGATTTTATTAATT 91913

RESULT 4
; Sequence 103, Application US/1121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR FILING DATE: 2004-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 103
; LENGTH: 160213
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

US-11-121-086-103

Query Match          0.3%; Score 21; DB 7; Length 160213;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2003 TATATTAAACAATCAGAGTA 2023
Db      159718 TATATTAAACAATCAGAGTA 159738

RESULT 5
; Sequence 28042, Application US/10750185
; Publication No. US2005026603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28042
; LENGTH: 991
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-28042

Query Match          0.3%; Score 20; DB 6; Length 991;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6190 CTATATTAATGTGAAAAA 6209
Db      340 CTATATTAATGTGAAAAA 321

RESULT 6
; Sequence 25079, Application US/10750185
; Publication No. US2005026603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25079
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-25079

Query Match          0.3%; Score 20; DB 6; Length 1077;
```

Best Local Similarity 100.0%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4961 AATGTTAAACAAATGAT 4980
DB 173 AATGTTAAACAAATGAT 192

RESULT 7
US-10-750-185-42078/c

; Sequence 42078, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42078
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Bovine 1986680617750
US-10-750-185-42078

Query Match 0.3%; Score 20; DB 6; Length 1206;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2196 ATTGTTTGAATAATTATT 2215
DB 133 ATTGTTTGAATAATTATT 114

RESULT 8

US-10-750-185-62576/c
; Sequence 62576, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62576
; LENGTH: 1324
; TYPE: DNA
; ORGANISM: Bovine 19866801499827
US-10-750-185-62576

Query Match 0.3%; Score 20; DB 6; Length 1324;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2441 TATATATTATTAAATATTC 2460
DB 173 TATATATTATTAAATATTC 154

RESULT 9
US-11-074-176-141/c

; Sequence 141, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAniff, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1362)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1401; NADH peroxidase/EC 1.11.1.1
US-11-074-176-141

Query Match 0.3%; Score 20; DB 7; Length 1362;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 497 AGATAAATTATTTATTCAT 516
DB 339 AGATAAATTATTTATTCAT 320

RESULT 10

US-10-750-185-56520
; Sequence 56520, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56520
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Bovine 19866801183934
US-10-750-185-56520

Query Match 0.3%; Score 20; DB 6; Length 1390;

Best Local Similarity 100.0%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2351 TAAATTTCTCAGATTTC 2370

DB 601 TAAATTTCTCAGATTTC 620

RESULT 11

US-10-750-185-35712/c
; Sequence 35712, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35712
; LENGTH: 1826
; TYPE: DNA
; ORGANISM: Bovine 1986680905746
US-10-750-185-35712

Query Match 0.3%; Score 20; DB 6; Length 1826;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2969 AAAATAATCTAATTAA 2988

DB 1423 AAAATAATCTAATTAA 1404

RESULT 12

US-11-112-908-32/c
; Sequence 32, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 32
; LENGTH: 193363
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-32

Query Match 0.3%; Score 20; DB 7; Length 193363;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3605 ATTTTCTTGCTTG 3624

DB 159346 ATTTTCTTGCTTG 159327

RESULT 13

US-10-928-446A-1/c
; Sequence 1, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; FILE REFERENCE: 0274-5785, 1US
; CURRENT APPLICATION NUMBER: US/10/928,446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
LOCATION: (825234)
OTHER INFORMATION: the 'n' at position 825234 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
LOCATION: (825270)
OTHER INFORMATION: the 'n' at position 825270 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
LOCATION: (825401)
OTHER INFORMATION: the 'n' at position 825401 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
LOCATION: (825428)
OTHER INFORMATION: the 'n' at position 825428 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
LOCATION: (825473)
OTHER INFORMATION: the 'n' at position 825473 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
LOCATION: (825765)
OTHER INFORMATION: a "c" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
LOCATION: (825828)
OTHER INFORMATION: the 'n' at position 825828 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
LOCATION: (826041)
OTHER INFORMATION: the 'n' at position 826041 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
LOCATION: (826546)
OTHER INFORMATION: the 'n' at position 826546 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
LOCATION: (826654)
OTHER INFORMATION: the 'n' at position 826654 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
LOCATION: (826826)
OTHER INFORMATION: the 'n' at position 826826 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
LOCATION: (826863)
OTHER INFORMATION: the 'n' at position 826863 may be 'a' or 'g'

```

FEATURE:
NAME/KEY: allele
LOCATION: (827008)
OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (843055)
OTHER INFORMATION: the 'n' at position 843055 may be 'g' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (843118)
OTHER INFORMATION: the 'n' at position 843118 may be 't' or 'c'
FEATURE:
NAME/KEY: allele
LOCATION: (871027)
OTHER INFORMATION: the 'n' at position 871027 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (871140)
OTHER INFORMATION: the 'n' at position 871140 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (871168)
OTHER INFORMATION: the 'n' at position 871168 may be 'c' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (872678)
OTHER INFORMATION: the 'n' at position 872678 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (872742)
OTHER INFORMATION: the 'n' at position 872742 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (925859)
OTHER INFORMATION: the 'n' at position 925859 may be 'c' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (993220)
OTHER INFORMATION: the 'n' at position 993220 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (993254)
OTHER INFORMATION: the 'n' at position 993254 may be 'g' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (1006462)
OTHER INFORMATION: the 'n' at position 1006462 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1007820)
OTHER INFORMATION: "ttct" may be deleted at this position
FEATURE:
NAME/KEY: allele
LOCATION: (1018038)
OTHER INFORMATION: the 'n' at position 1018038 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1018704)
OTHER INFORMATION: the 'n' at position 1018704 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1018718)
OTHER INFORMATION: "gtc" may be deleted at this position
FEATURE:
NAME/KEY: allele
LOCATION: (1026786)
OTHER INFORMATION: the 'n' at position 1026786 may be 'c' or 'a'
FEATURE:
NAME/KEY: allele
LOCATION: (1047134)
OTHER INFORMATION: the 'n' at position 1047134 may be 'a' or 'g'
FEATURE:

```

```

NAME/KEY: allele
LOCATION: (1047159)
OTHER INFORMATION: the 'n' at position 1047159 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1047378)
OTHER INFORMATION: the 'n' at position 1047378 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1047739)
OTHER INFORMATION: the 'n' at position 1047739 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1050133)
OTHER INFORMATION: "ttaaa" may be deleted at this position
FEATURE:
NAME/KEY: allele
LOCATION: (1050539)
OTHER INFORMATION: the 'n' at position 1050539 may be 'c' or 't'
FEATURE:
NAME/KEY: allele
LOCATION: (1062808)
OTHER INFORMATION: the 'n' at position 1062808 may be 'c' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1066392)
OTHER INFORMATION: the 'n' at position 1066392 may be 'a' or 'g'
FEATURE:
NAME/KEY: allele
LOCATION: (1073711)
OTHER INFORMATION: the 'n' at position 1073711 may be 'c' or 't'
US-10-928-446A-1

```

```

Query Match      0.3%; Score 20; DB 6; Length 1080000;
Best Local Similarity 100.0%; Pred.No. 69;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      5424 ATTATGTTTGTGACTCTGA 5443
Db      244602 ATTATGTTTGTGACTCTGA 244583

```

```

RESULT 14
US-10-928-446A-181/c
; Sequence 181, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; TITLE OF INVENTION: VIRAL BUDGING
; FILE REFERENCE: 0274-5785.1US
; CURRENT APPLICATION NUMBER: US/10/928.446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 181
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (826985)
OTHER INFORMATION: full exon 1 range is 826667-827008
FEATURE:
NAME/KEY: allele
LOCATION: (827008)
OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
FEATURE:
NAME/KEY: CDS
LOCATION: (843242)
OTHER INFORMATION:

```



```
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (922549)..(922630)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (926021)..(926059)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (929123)..(929176)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (993104)..(993154)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (999547)..(999608)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1000354)..(1000456)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1002118)..(1002284)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1006117)..(1006249)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1007860)..(1008036)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1010940)..(1011014)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1018160)..(1018291)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1018800)..(1018919)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1020028)..(1020225)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1026559)..(1026736)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1028113)..(1028167)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1034316)..(1034374)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1041390)..(1041455)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1043121)..(1043350)
OTHER INFORMATION: exon
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: (1044868)..(1044989)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1047519)..(1047589)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1050296)..(1050391)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1062648)..(1062708)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1060368)..(1060441)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1064561)..(1064620)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1066207)..(1066314)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1067768)..(1067864)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1068609)..(1068681)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1073289)..(1073388)
OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-181

Query Match          0.3%; Score 20; DB 6; Length 1080000;
Best local Similarity 100.0%; Pred. No. 69;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5424 ATTATGTTTGTGACTCTGA 5443
Db       244602 ATTATGTTTGTGACTCTGA 244583

RESULT 15
US-10-928-446A-183/c
; Sequence 183, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF MEDD4L ASSOCIATED WITH HYPERTENSION AND
; FILE REFERENCE: 0274-5785.1US
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 183
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

```

? LOCATION: (826985) .. (827017)
? FEATURE:
? OTHER INFORMATION: full exon 1 range is 82667-827008
? NAME/KEY: CDS
? LOCATION: (843242) .. (843315)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (922549) .. (922630)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (926021) .. (926059)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (929123) .. (929176)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (993104) .. (993154)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (999547) .. (999608)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1000354) .. (1000456)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1002118) .. (1002284)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1006117) .. (1006249)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1007860) .. (1008036)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1010940) .. (1011014)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1018160) .. (1018291)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1018800) .. (1018919)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1020028) .. (1020225)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1026659) .. (1026736)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1028113) .. (1028167)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1034316) .. (1034374)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS

```

```

? LOCATION: (1041390) .. (1041455)
? FEATURE:
? OTHER INFORMATION: exon
? NAME/KEY: CDS
? LOCATION: (1043121) .. (1043350)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1044868) .. (1044989)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1047519) .. (1047589)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1050296) .. (1050391)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1060368) .. (1060441)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1062648) .. (1062708)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1064561) .. (1064620)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1066207) .. (1066314)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1067768) .. (1067864)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1068609) .. (1068681)
? OTHER INFORMATION: exon
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1073289) .. (1073388)
? OTHER INFORMATION: full exon 30 range is 1073289-1075279
? FEATURE:
? OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-183

```

```

Query Match 0.3%; Score 20; DB 6; Length 1080000;
Best Local Similarity 100.0%; Pred.No. 69;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5424 ATTATGTTGTGACTCTGA 5443
DB 244602 ATTATGTTGTGACTCTGA 244583

```

Search completed: December 26, 2005, 06:19:55
Job time : 618 secs

THIS PAGE BLANK (USPTO)